Journal of Applied Psychology

Edited by Donald G. Paterson, University of Minnesota

Consulting Editors

George K. Bennett, *Psychological Corporation*
Harold E. Burtt, *Ohio State University*
Allen L. Edwards, *University of Washington*
Clifford E. Jurgensen, *Minneapolis Gas Co.*
Irving Lorge, *T. C. Columbia University*
Quinn McNemar, *Stanford University*
Alexander Mintz, *City College of New York*
James P. Porter, *Danville, Illinois*
Harold F. Rothe, *Amer. Hosp. Supply Corp., Evanston, Ill.*
Julian B. Rotter, *Ohio State University*
Edward K. Strong, Jr., *Stanford University*
Donald E. Super, *T. C. Columbia University*
Morris S. Viteles, *University of Pennsylvania*
Alfred C. Welch, *Knox-Reeves, Minneapolis*

Table of Contents

An Opinion Survey of a Regional Union Group: K. Davis and E. E. St. Germain 285
Attitudes of Personnel Managers and Student Groups toward Labor Relations: R. E. Bernberg 291
Empathy Test Scores of Union Leaders: R. H. Van Zelst 293
An Appraisal of Worker Characteristics as Related to Age: W. H. Bowers 296
Relation between How Supervised, Intelligence, and Education for a Group of Supervisory Candidates in Industry: F. R. Wickert 301
Changes Occurring in Teacher-Pupil Attitudes During a Two-Weeks Guidance Workshop: J. Shaw, H. J. Klausmeier, A. H. Luker, and H. T. Reid 304
Personality Characteristics Related to Success in Practice Teaching: H. G. Gough and W. H. Pemberton 307
Predictive Value of The Empathy Test in Automobile Salesmanship: F. P. Tobolski and W. A. Kerr 310
Predicting Success of Students in Veterinary Medicine: W. L. Layton 312
The Criterion Problem in the Prediction of Medical School Success: E. L. Gaier 316
A Comparison of Three Criteria of Aircrew Effectiveness in Combat Over Korea: J. K. Hemphill and L. B. Sechrest 323
Employment Prognosis of the Post-Poliomyelitic: L. V. Wendland 328
A Job Preference Survey for Industrial Applicants: W. F. Long 333
Development of a Home Economics Interest Inventory: H. Johnson 338
Effect of Changing the Number of Item Responses from Five to Four in the Same Test: A. Swordes 342
Some Research Findings with the Wonderlic Personnel Test: E. N. Hay 344
Speed and Accuracy of Reading Arabic and Roman Numerals: D. K. Perry 346
The Dimensional Analysis of Motion: IV. Transfer Effects and Direction of Movement: P. Von Trebra and K. U. Smith 348
Book Reviews 354
New Books, Monographs, and Pamphlets 359

American Psychological Association

Vol. 36, No. 5 October, 1952

Journal of Applied Psychology

Published Bi-monthly by the American Psychological Association, Inc.
Prince and Lemon Sts., Lancaster, Pa.

Annual subscription, $6.00; single copies, $1.25

Subscriptions and business communications should be sent to

American Psychological Association
1515 Massachusetts Avenue N.W.
Washington 5, D. C.

Articles for publication and books for review should be sent to the Editor

Professor Donald G. Paterson, Department of Psychology
University of Minnesota, Minneapolis 14, Minnesota

This Journal gives prompt consideration to manuscripts reporting *original investigations* in any field of applied psychology except clinical and consulting psychology. A descriptive or theoretical article is occasionally accepted if it deals in a distinctive manner with a problem of applied psychology. The policy is, however, to favor papers dealing with quantitative investigations of direct value to psychologists working in the following fields: Vocational diagnosis and occupational guidance; educational diagnosis, prediction and guidance at the secondary school level and higher; personnel selection, training, placement, transfer and promotion in business, industry and government service including the armed forces; supervisory training in business, industry and government; bio-mechanics or design of machines to fit the human operator; illumination, ventilation and fatigue in industry; job analysis, description, classification and evaluation; measurement of morale of executives, supervisors, or employees; surveys of opinion on social or political issues, such as those conducted by The Psychological Corporation; psychological problems in market research and in advertising.

Articles may be under 500 words. The maximum is 12,000 words, the average in the neighborhood of 4,000 words. To reduce lag of publication, adherence to the rule of "brevity consistent with clarity" is encouraged.

A lapse of six to twelve months occurs between acceptance of an article and its publication, the lag varying with the rate at which manuscripts are submitted. If, however, an author is prepared to defray the costs of printing the necessary extra pages, he may arrange for earlier publication without thereby postponing the appearance of manuscripts by other contributors. This enables the management to provide space in addition to the scheduled 64 pages per issue. "Early publication" is thus a direct contribution to the subscribers. By cutting down lag in publication, it also benefits those authors whose articles are published in regular turn.

Tables, footnotes and references as well as text of manuscripts should be typed double-spaced throughout. Authors should adhere to the conventions described by J. E. Anderson and W. L. Valentine in "The preparation of articles for publication in the journals of the American Psychological Association," *Psychol. Bull.*, 1944, 41, 345–376. A reprint of this article will be loaned to any prospective contributor who does not find it in his library.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the act of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized October 10, 1947

Copyright, 1952, by The American Psychological Association, Inc.

Journal of Applied Psychology

VOL. 36, NO. 5 OCTOBER, 1952

An Opinion Survey of a Regional Union Group

Keith Davis[1]
Indiana University
and
Edward E. St. Germain
San Francisco Port of Embarkation

Although employee opinion surveys are widely accepted and used in the United States today,[2] survey variations have been generally omitted by those making them. One variation seldom used is to have the union organization sponsor the survey and distribute the survey questionnaire. Another variation is to have the union administration use the survey to determine its members' attitudes toward its effectiveness in doing its own job, as well as to determine their attitudes toward their employer. This article reports the results of an opinion survey of 140 members of a midwestern union, early in 1951, using the variations described.

If one excludes opinion surveys of workers for public reporting about employers and unions "in general," one finds that the usual survey practice is for the employer to administer the questionnaire to his employees or to have a consultant administer it. Unions have many times "cooperated with" an employee opinion poll or occasionally co-sponsored a poll, but generally they have not independently sponsored opinion surveys of their members. At the time this survey was made, the research departments of both C I O and A F of L national headquarters informed the authors that they had no knowledge of any union having conducted a formal poll of its members' attitudes toward company management practices. Use of the union as survey sponsor probably gives the employee added confidence that his comments about management will not be used against him by management. This should encourage him to give more accurate responses. On the other hand, it is conceivable that he might attach to the survey a union-centered antagonism toward management. This would cause a union-sponsored survey to be biased against management.

If a union sponsors an attitude survey about management, it seems appropriate to use the same survey to determine how its members feel about its own administration. Attitude surveys are not just for business. The union, too, is an organization with a management hierarchy and management problems. It performs all of the administrative functions performed by business management, though sometimes in a different way. It works with the same human raw material that business management does. It, therefore, has the same obligations that other managements have to use human tools (like the attitude survey) to help bring human problems to a human solution.

Description of the Group

The area of the survey was an A F of L union local situated in a midwestern state. The local headquarters was in an industrial city

[1] Mr. Davis is Associate Professor of Management, Indiana University, and Mr. St. Germain is Organization and Methods Examiner, San Francisco Port of Embarkation. This research was completed while the authors were Associate Professor of Industrial Relations and Research Associate, respectively, at the University of Texas.

[2] For example, a recent report states that 35 per cent of 160 employers have used employee opinion surveys and that over 50 per cent of those not using them believe them of some value. See Shurtleff, W. E. Is management listening? *Personnel*, 1951, 27, 101–102.

of over 100,000 population which had a climate considered "favorable" to unionism. The 1,200 members operated many varieties of heavy machinery used in commercial construction and road building, such as draglines, bulldozers, concrete mixers, pile drivers, cranes, and tractors. The membership at the time of the survey was estimated to be 60 per cent skilled, 30 per cent semi-skilled, and 10 per cent unskilled, but it varied somewhat according to seasons and work demands. Hourly wage rates ranged from about $1.60 to $2.35.

Since the union members work on construction jobs, most of them work in small groups of two to twenty-five in different towns and rural locations throughout the state. Each member regularly works with various employers over a period of time, remaining with one employer only as long as the type of equipment he operates is needed. Because of the wide geographical dispersion and movement of membership, the union maintains between five and ten area and branch offices in other cities.

The key union administrative official is the full-time business agent, whose office is in the local headquarters. Employers list orders for operators with him, and he assigns operators—and removes them—according to his judgment. He is an elected officer and has been in office over ten years.

The local members generally have more contact with the union than they do with any individual employer, because they work for different employers, but are always assigned and removed by the same union. Even on the job, the members are largely free of regular contact with their employer and free of close supervision by him. The union does not have a union shop in the industry, but the non-union group of similar workers is very small.

Survey Method

The business agent agreed to sponsor the survey, after the authors suggested it. The agent helped the authors prepare a three-page questionnaire of 28 questions. Each question was short, simple, and pitched to the readability of the respondents, who answered by checking one of five responses on an attitude scale. One of the available responses was "undecided." The questionnaire heading was "*Wanted: Your Opinion of Your Union and Your Job.*" The written directions were a part of the survey; they assured the employee that the survey was anonymous and that the results would be available to him through the union. Prior to answering the survey, each employee answered general questions about such subjects as his age group and marital status.

The questionnaire was pretested on a few union workers not in this union. Revised questionnaires (301 in number) were then sent to the business agent for distribution. A few days previously he had notified the membership that the survey was pending. Since the potential respondents were working in various locations over the state, the agent distributed part of the questionnaires personally during his regular tour to different jobs. The remainder of the questionnaires were distributed through members of the Executive Board who met at headquarters during the survey period. Reliability of the returns was checked by the cumulative frequency method,[3] which showed no important difference in replies resulting from the two methods of distribution.

Replies received numbered 151 out of the 301 sent. Of those received, 11 were excluded because of incompleteness, leaving 140 replies for analysis. These 140 replies represented 47 per cent of questionnaires distributed and about 12 per cent of all dues-paying members. Replies were received from about 45 per cent of the active job locations in the state. Respondents readily admitted they *had opinions*, because of the fact that the "undecided" choice received no more than five per cent of responses to any question.

Results

Attitudes Toward Union Administrative Practices and Problems. Union members generally approved of their local's administrative practices because their "favorable" responses at no time dropped below 70 per cent of total responses (see Table 1). However, in most instances a greater proportion was "moderately favorable" rather than "very favorable." The weakest part of union administration, from

[3] L. O. Brown. *Marketing and distribution research.* The Ronald Press Company, 1949, pp. 524–528.

Table 1

Union Member Opinions of Selected Union Administrative Practices and Problems, Arranged in Descending Order of Favorable Responses Based on 140 Replies

Subject of Opinion	Per Cent of Responses						
	Total Favorable Responses	Total Unfavorable Responses	Most Favorable Response	Second Most Favorable Response	Second Most Unfavorable Response	Most Unfavorable Response	Undecided
Worthwhileness of union meetings	90	9	48	42	6	3	1
Likelihood that present contract will be renewed without major changes	86	13	19	67	9	4	1
Administration of grievances	84	11	32	52	9	2	5
Desire to attend union meetings	83	16	56	27	13	3	1
Probability that members of the local personally support same political candidates	80	17	46	34	10	7	3
Local's comparative strength in bargaining	77	22	24	53	15	7	1
How well respondent personally understands labor contract	77	23	19	58	23	0	0
Extent to which local keeps its members informed on union matters	76	21	22	54	14	7	3
Extent to which union members in general are in the "right" when they call a strike	76	22	30	46	19	3	2
Opinion of present labor contract	73	27	20	53	21	6	0
Local's fairness in assigning jobs	70	26	24	46	16	10	4

members' viewpoints, was the business agent's fairness in assigning jobs. Only 70 per cent felt that he did not show favoritism. The strongest part of union administration was its conduct of meetings. Ninety per cent felt that meetings were worth the time spent, and 83 per cent desired to attend union meetings "usually" or oftener. It should be noted that far less than 83 per cent actually attended meetings, because most meetings were held in union headquarters and members had to travel a considerable distance to them.

Only 20 per cent of the members felt that the current bargaining contract was "excellent," but another 53 per cent rated it "good." The remaining 27 per cent had a low opinion of the contract, which indicates that more than one out of four members will want contract improvements at the next contract period. In fact, 13 per cent believed that the contract would not be renewed again without major changes. It is interesting that the proportion who thought the labor contract was excellent (20 per cent) was about the same as the proportion who thought they understood all of the contract.[4] This suggests a relation between degree of understanding of the contract and satisfaction with it. Further, 23 per cent had "little" understanding of the contract and this indicates that much more contract communication needs to take place. Twenty-one per cent felt that they were not well informed on union matters in general.

A significant factor in the employees' over-all viewpoints may be that 77 per cent believed their union was "usually" or "always" stronger than the employers at the bargaining table. In

[4] This suggests the desirability of simplified language in union contracts. See Lauer, Jeanne, and Paterson, D. G. Readability of union contracts. *Personnel*, 1951, **28**, 36–40; and Tiffin, J., and Walsh, F. X. Readability of union-management agreements. *Personnel Psychol.*, 1951, **4**, 327–338.

other words, they could "afford" to want more at the bargaining table, because they had the strength to get it. There is some logical basis for their opinions, since the employer group consisted of many independent employers, but the union group operated as one organized unit. Although the union and management have worked together without strike for over a decade, the union's confidence in its power would undoubtedly affect its conduct during a severe strike. From another viewpoint, employer conduct at each bargaining session may be influenced by the fact that union members have so much confidence in their own strength.

Eighty-four per cent of the respondents agreed that the union's handling of grievances was satisfactory, but only 32 per cent would go so far as to say that grievances were "always" handled satisfactorily. Inasmuch as management and the union have amicably settled grievances without resort to strike during the last ten years, one might wonder why more members were not strongly satisfied with grievance administration. One explanation is that grievance settlements usually involve compromise, which at best leaves the employee moderately satisfied, but not always strongly satisfied.

Most members (eighty per cent) believed that they usually supported the same political candidates, which indicates that their leadership has successfully created within their union environment a feeling of political solidarity. As a corollary question, members were asked if they thought labor unions have a strong influence in politics, and 84 per cent responded affirmatively. Furthermore, 76 per cent felt that unions belonged in politics.[5] It seems clear that the respondents favored union political action, felt that their associates were cooperating to support the same candidates, and felt that they strongly influenced election results.

A general question was asked: "How often are labor unions in the right when they call a strike?" Forty-six per cent checked "usually" and thirty per cent checked "always" which indicates a high degree of belief in their cause and in the strike as a device of coercion.

Attitudes Toward Employer Administrative Practices and Working Conditions. Attitudes toward the employer group were not as consistent as those toward the union, but they were over-all about as favorable. Favorable responses did not drop below 66 per cent for any subject (see Table 2). The highest proportion of favorable responses during the entire survey resulted from the question, "How often do the contractors interfere with your job performance?" Ninety-three per cent of the respondents believed that employers usually did not interfere with job performance. This proportion speaks well of employer supervision and employee ability to do the job without close supervision.

Respondents were moderately satisfied with their working conditions. Seventeen per cent of the 140 respondents felt that their job was never monotonous, and another 71 per cent felt that their job was only occasionally monotonous. The respondents' jobs are generally viewed as non-monotonous by their employers, who point out that they work independently or in small groups and are to a considerable extent "their own boss" on the job. Another factor relieving monotony is the constant change of employers. Seventy-one per cent approved their present practice by saying that they did not desire to work for the same employer regularly.

Seventy-five per cent believed that most employers furnished good tools and equipment, but 24 per cent thought few employers did. Although equipment in this industry is expensive and recently has been difficult to purchase, employers should through good communication and equipment maintenance be able to reduce this strong minority that is dissatisfied with equipment. Respondents rated their over-all working conditions almost identically with their rating of equipment. Seventy-six per cent believed conditions were good or excellent, but a strong minority of 21 per cent thought working conditions were only "fair." Considering the fact that much of their work is rough, dangerous, outside work away from home with different employers, the proportion of satisfac-

[5] In a national survey in November, 1947, only 59 per cent of the respondents believed unions should participate in politics. See What the factory worker really thinks. *Factory Mgmt. & Maint.*, November, 1947, p. 93.

Table 2

Union Member Opinions of Selected Employer Administrative Practices and Job Working Conditions, Arranged in Descending Order of Favorable Responses Based on 140 Replies

Subject of Opinion	Per Cent of Responses						
	Total Favorable Responses	Total Unfavorable Responses	Most Favorable Response	Second Most Favorable Response	Second Most Unfavorable Response	Most Unfavorable Response	Undecided
Extent to which employer does not interfere with job performance	93	7	29	64	6	1	0
Freedom of job from monotony	88	12	17	71	8	4	0
Degree of employer cooperation with local	81	18	12	69	14	4	1
Over-all opinion of present job compared to factory work at same pay	80	18	64	16	12	6	2
Over-all appraisal of working conditions	76	23	9	67	21	2	1
Extent to which good tools and equipment are provided	75	25	9	66	24	1	0
Over-all attitude toward employers	69	29	12	57	24	5	2
"Present wage scale is as high as it should be"	67	28	7	60	23	5	5
Over-all appraisal of morale of *other* local members	66	31	12	54	23	8	3
Desire to work for same employer continuously	26	71	6	21	55	16	2

tion with working conditions seems to be reasonable.

Employees were least satisfied with their wages. Twenty-eight per cent disagreed with the statement, "Your present wage scale is as high as it should be." A negative phrasing of this controversial statement might have produced different results.

More than a majority of the respondents had a favorable over-all opinion of employers and believed they were positively trying to cooperate with the union. These opinions reflect a decade of harmonious relations between employer and employee. Eighty-one per cent believed that employers usually try to cooperate with their union. Sixty-nine per cent rated their employers "good" or "excellent" as employers. Since each respondent had worked for many employers, the sum of the 140 respondents' judgments should include almost every employer in the group. Employees' appraisal of their employers was almost identical to their appraisal of the morale of *other* employees, which indicates that they thought others felt just as they did about their mutual jobs and bosses.

The problem of morale and working conditions was approached from another point by asking employees, "Would you change places with a factory worker at the same pay?" Sixty-four per cent said "No" and another 16 per cent said "Unlikely," which is further evidence of their general job satisfaction.

Summary

It is practical and useful for a union organization to sponsor an employee attitude survey, in the same way that attitude surveys by employers have proven useful. Furthermore, it seems appropriate that the survey be directed toward union practices as well as employer practices. Both union and employer share

authority in determining the policy under which the worker works. Both are in frequent contact with him and affect his work situation. There is no reason to study his attitude toward only the employer.

Both the union and the employers developed an early interest in the survey reported herein, even though only the union sponsored it. Both groups felt that its results would help them understand the strengths and weaknesses of their relationship.

The survey showed that employees were reasonably satisfied with their union and their employers, which reflected many years of cooperation. However, an important minority ranging from 21 to 27 per cent were not satisfied with the following union practices: fairness in assigning jobs, provisions of the present labor contract, and the local's communications on union matters, including the labor contract. Members were moderately satisfied with grievance administration, only 11 per cent dissenting. They also were strongly satisfied with union meetings. Furthermore, three-fourths (77 per cent) believed they were stronger than employers at the bargaining table. They also favored union political action (76 per cent), felt that their associates were cooperating to support the same political candidates (80 per cent), and felt they strongly influenced election results (84 per cent).

Most employees were satisfied with their job freedom (93 per cent), lack of job monotony (88 per cent), degree of employer cooperation with their local (81 per cent), desirability of their job compared to a factory job (80 per cent), and over-all working conditions (76 per cent). However, a strong minority felt that the attitudes of other employees toward employers were only fair or poor (31 per cent), which indicates a need for better understanding between employees. A sizeable minority also were dissatisfied with wages (28 per cent), with the necessity to work for different employers (27 per cent), and with the quality of tools and equipment furnished by the employer (25 per cent).

The main dissatisfactions uncovered in this survey were in the areas of communication and practical day-to-day problems, rather than in areas producing a conflict of principle. The day-to-day problems can be resolved or reduced within the existing pattern of employer-union cooperation, aided by such techniques as union-sponsored and management-sponsored attitude surveys.

Received November 26, 1951.

Attitudes of Personnel Managers and Student Groups toward Labor Relations

Raymond E. Bernberg

Los Angeles State College

In pursuing a college course of study in preparation for a career in personnel work in Industry, a student may choose between two possible major studies: Industrial Psychology (in the Department of Psychology), Personnel Management (in the Department of Business).

It seemed to the author that a dichotomy in attitudes and values exists between these two groups of students, both of whom are preparing for the same career. This suggested three problems for investigation: 1) Can this apparent difference in attitudes and values be defined by psychological tools? 2) Does a similar dichotomy exist among personnel managers? 3) Which student group is more similar in attitudes and values to personnel managers?

Procedure

The Allport-Vernon Study of Values (**1**) and the Labor-Relations Information Inventory (**4**)[1] were selected as instruments to measure the differences between the groups. The Allport-Vernon scale of values measures relative strength of six basic interests or motives: *theoretical*, *economic*, *aesthetic*, *social*, *political*, and *religious*. The Labor-Relations Inventory indirectly assesses attitudes toward labor.

The tests were administered to 51 Industrial Psychology majors who were upper-division and graduate students from colleges and universities in the Los Angeles area and from two universities in the middle-west; 61 Personnel Management majors from the same sources; and 43 Personnel and Employment Managers from the Los Angeles area. The subjects constitute an approximate 30% of those to whom materials were given. The whole sample is dominantly male with but a few female subjects throughout.

[1] To reduce printing costs the Labor-Relations Information Inventory has been deposited with the American Documentation Institute. Order Document 3643 from American Documentation Institute, 1719 N Street N.W., Washington 6, D. C., remitting $1.00 for microfilm (images 1 inch high on standard 35-mm. motion-picture film) or $1.00 for photocopies (6 × 8 inches) readable without optical aid.

Results

The results of an analysis of the data are contained in Table 1. The students as a group have significantly higher mean scores in their attitude toward labor than those in the field. According to the norms of the test, the student means lie in the "neutral zone" of attitude toward labor, while those in the business world have a mean score which is in the "pro-management zone" of attitude toward labor. The employed group also have a narrower range of scores, indicating greater homogeneity. Both student groups have broader ranges, reflecting their heterogeneity with respect to this attitude.

On the Scale of Values test, Industrial Psychology majors have significantly higher mean scores for the *theoretical* and the *aesthetic* values than either of the other two groups. They have significantly lower mean scores than the other two groups on the *economic* and on the *religious* values. The Personnel Management students and the Personnel and Employment Managers show similar means on all of the values.

In interpreting these data, one should recognize the great differences in training between the Industrial Psychology and Personnel Management students. The high value of the one for the *theoretical* opposed to the low value for the *economic* may be an indication of the emphasis in curriculum and course material in psychology, even in the applied area. Nevertheless, two possibilities suggest themselves: (1) students electing the Industrial Psychology major may have entered college with a more liberal attitude toward labor and a more theoretical orientation; (2) the differences appearing between the two groups may

Table 1

Mean Scores, Standard Deviations, and Tests of Significance for Students Majoring in Industrial Psychology, Students Majoring in Personnel Management and for Employed Personnel Managers on the Allport-Vernon Scale of Values Test and the Indirect Test of Attitude Toward Labor

			Allport-Vernon Scale of Values					
Group*		Labor Relations Inventory	Theoretical	Economic	Aesthetic	Social	Political	Religious
A	Mean	13.56	35.6	30.6	27.9	30.8	30.8	24.5
	S.D.	4.70	7.58	7.07	8.39	6.37	6.27	7.98
B	Mean	12.93	29.6	35.1	22.7	30.0	31.9	31.3
	S.D.	4.68	6.18	7.11	6.29	5.27	5.40	8.09
C	Mean	9.26	30.5	33.7	22.2	30.6	32.6	30.0
	S.D.	4.44	5.86	6.32	7.06	4.27	4.44	8.53
		C.R. P.	C.R. P.	C.R. P.	C.R. P.	C.R. P.	C.R. P.	C.R. P.
A–B		.71 .48	4.60 .001	3.42 .001	3.75 .001	.73 .46	1.10 .27	4.52 .001
A–C		4.53 .001	3.65 .001	2.24 .025	3.56 .001	.18 .86	1.63 .11	3.24 .002
B–C		4.06 .001	.74 .46	1.07 .29	.36 .72	.63 .53	.71 .48	.76 .45

* Group A—Industrial Psychology Majors (N = 51); Group B—Personnel Management Majors (N = 61); Group C—Personnel and Employment Managers (N = 43).

result from differences in expressed attitudes of their instructors within the different curricula (**2, 3**).

In other words, are attitudes and values such as we are discussing engendered by training in appropriate curricula, or is basic motivation reflected in these attitudes and values and manifested in the choice of the different courses of study? It is quite possible, of course, for these two factors to coexist.

If it is true that the attitudes and values are stable and resistant to change, it seems likely that most students in the Industrial Psychology major face an experience of frustration, through inability to accept the attitudes and values of those under whose executive direction they will work. We do not know the answer to this problem, but it is evident that the answer is needed to give us a better basis for vocational guidance in this area.

Received November 24, 1951.

References

1. Allport, G. W., and Vernon, P. E. A test for personal values. *J. abnorm. soc. Psychol.*, 1931, **26**, 231–248.
2. Seashore, H. G. Validation of the *Study of Values* for two vocational groups at the college level. *Educ. psychol. Measmt*, 1947, **31**, 249–259.
3. Stagner, R. *Psychology of personality*. New York: McGraw-Hill Co., 1949.
4. Weschler, I. R. An investigation of attitudes toward labor and management by means of the error-choice method. *J. soc. Psychol.*, 1950, **32**, 51–70.

Empathy Test Scores of Union Leaders

Raymond H. Van Zelst

Illinois Institute of Technology

A careful survey of recent psychological literature reveals numerous articles relating to or a component of the concept of empathy. However, in extricating the kernel of empathy from the abundance of associative material one becomes aware of a rather marked lack of concerted effort dealing with the direct measurement of this important talent. In no way is this ablation of empathy meant to deter from the important, and, in fact, corporate factors (e.g., sympathy, insight, etc.) contributing to the concept of empathy. Rather, the author herein seeks to adopt an isolative view of the problem by dealing with a concept which might be most appropriately entitled *massempathy*.

The particular trait known as *empathy* has been defined by various sources (**5**, **9**, **10**, **11**, **12**, **13**) as the "ability to put oneself in the other person's position, establish rapport, and anticipate his reactions, feelings, and behaviors." The concept of *massempathy* for which the instrument used in this study is designed may be said to connote the ability of an individual to "feel into" the average or hypothetically average person's situation and so react (**11**). This trait may be seen then as being empathic ability plus, since one must not only "feel into" or interject oneself into the structural field of another specific person, but also assume the "average" person's structural field. "The difference between our official or average or cultural expectation of action in a situation and the actual conduct of the person indicates the presence of a private interpretation. Where the private version equals the official one we have normal conduct" (**4**).

This article, therefore, will attempt to present evidence pertaining to the validity and reliability of a recently devised test designed to assess empathic or massempathic ability.

Description of the Test

In accord with the above definition of empathy it was the belief of the authors of the *Empathy Test* that "individuals who are superior in empathic ability are persons who are above average in understanding and anticipating the reactions of other people" (**8**). This belief was utilized in the construction of the test.

The *Empathy Test* is composed of three parts. Part I consists of various types of music (e.g., polkas, classical, etc.) and the subject is requested to "rank these types in order of their probable popularity among the non-office factory workers of the United States." The types of music used are based upon the advice of an RCA Victor phonograph records sales executive and the normative data supporting this section are the results of a national survey program.

Part II contains the titles of 15 current well-known magazines and the testee is required to rank them "in order from most to least total paid circulation." Actual rankings were obtained from published actual sales of the various publications.

Part III has the subject rank ten commonly annoying experiences according to their specific annoyance to the general population. Normative reports are based upon the extensive findings of Hulsey Cason (**2**).

Each section, therefore, requires the individual to define specific normative reactions and so is in keeping with the proposed definition of empathy. Scoring of the test is objective and the amount of time needed for administration short (15 min.).

Subjects and Criteria

Subjects used in this validation of the *Empathy Test* are 64 business agents—a total sample of five A F of L building trades unions situated in the Chicago area. The main duties of these men are to: (1) enforce union rules; (2) recruit and organize new members; (3) settle grievances and disputes; and (4) handle all benefit payments. All of these duties may be considered as constituting activities which

require definite leadership abilities such as tact, justice, sound judgement, determination, responsibility, etc. for efficient performance (3).

Six criteria were selected as gauges for the leadership ability of these business representatives. The first is a ranking of these men as to "Success at present leadership level and probability of leadership ability at a higher leadership level" by their immediate district superiors. Each of the five groups was ranked by four judges, the district president, vice-president, secretary, and treasurer of the particular union, working independently. All of the judges were well acquainted with each individual's record as a union official. Reliabilities of the rankings as determined by the method of average intercorrelation (7) were .92, .91, .91, .89, and .88.

The second criterion utilized was voting record, that is, the percentage of the total number of votes cast which the individual received in the most recent local election. The validity of this criterion may be contaminated by the caliber of the opposition.

The third criterion was score on the File-Remmers *How Supervise?* test (Form M), a reliable and validated measure of supervisory knowledge and insight (1).

The fourth criterion used was the recruitment and organization of new members. The main contamination present here is the fertility of the area to which the operations of the business agents are limited. This was controlled by taking per cent increase over previous year's local size.

The fifth criterion employed was the ability to settle grievances and disputes. Complete records are kept by the district offices of all such occurrences which require the attention of the local business agent. These records were taken individually and carefully evaluated in detail by the district president, vice-president, secretary, and treasurer, working individually, and assigned a point rating of 0, 3, 6, 9, or 12 according to estimated successfulness of handling. The mean accumulated score was used as the index of proficiency.

The sixth criterion utilized was the enforcement of rules and regulations. Violations of union working rules other than grievances and disputes which were reported by sources other than the local unions' business representative were used as an index of laxity. In order to minimize possible errors due to the opportunity for such violations to arise, the number of violations reported was divided by the number of jobs in progress in the area of the local representative's jurisdiction.

The author then endeavored to ascertain the success of these business agents in this complex activity by singling out these six applications of leadership ability and measuring them. These six aspects are in the opinion of the cooperating district officials the keystones of the entire activity of these union business agents. Furthermore, these significant criteria are readily observed and accurately measured.

Results

Pearsonian correlation coefficients were computed between *Empathy Test* score and the six criteria. The business representatives' score on the *Empathy Test* correlated .67 with leadership rank, .38 with percentage of vote received in local union election, .55 with *How Supervise?* score, .60 with recruitment and organization of new members, .64 with the ability to settle grievances and disputes and .44 with enforcement of rules and regulations. The multiple R between *Empathy Test* score and the six criteria is .76. All coefficients are statistically significant at the one per cent level of confidence.

Reliability of the *Empathy Test* was ascertained by means of the odd-even technique. Odd items on each subtest were utilized as one score while the even numbered items constituted the other test score. Corrected reliability was .87.

The substantial coefficients arrived at between test score and the six criteria of leadership ability seem to suggest that the *Empathy Test* may be profitably employed in the prediction and selection of potential union leaders. It is evident, however, that the findings need to be cross-validated before they may be accepted as fully established.

Received October 17, 1951.

References

1. Bruce, M. E. *The third mental measurements yearbook.* New Brunswick: Rutgers Press, 1949.
2. Cason, H. *Common annoyances: a psychological study of everyday aversions and irritations.*

Psychol. Monogr., 1930, **40**, No. 2 (Whole No. 182).
3. Coffin, T. E. A three component theory of leadership. *J. of abnorm. soc. Psychol.*, 1944, **39**, 63–83.
4. Dollard, J. *Criteria for the life history.* New Haven: Yale Press, 1935.
5. Dymond, R. F. A preliminary investigation of the relation of insight and empathy. *J. consult. Psychol.*, 1948, **12**, 228–233.
6. Dymond, R. F. A scale for the measurement of empathic ability. *J. consult. Psychol.*, 1949, **13**, 127–133.
7. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1937.
8. Kerr, W. A., and Speroff, B. J. *Measurement of empathy.* Box 1625, Chicago 90: Psychometric Affiliates, 1951.
9. Miller, F. G., and Remmers, H. H. Studies in industrial empathy: II. Management's attitude toward industrial supervisors and their estimates of labor's attitude. *Personnel Psychol.*, 1950, **4**, 33–40.
10. Remmers, L., and Remmers, H. H. Studies in industrial empathy: I. Labor leaders attitude toward industrial supervisors and their estimates of labor's attitude. *Personnel Psychol.*, 1949, **3**, 427–436.
11. Speroff, B. J. *Evaluation of a mass empathy measurement on male college students.* Paper delivered at Midwest. Psychological Association meeting, Chicago, 1951.
12. Van Zelst, R. H. Validation evidence on the empathy test. *Educ. psychol. Measmt*, in press.
13. Warren, H. C. *Dictionary of psychology.* Boston: Houghton Mifflin Co., 1934.

An Appraisal of Worker Characteristics as Related to Age*

William H. Bowers

The Ohio State University

Since 1900 the number of persons age 65 and over has grown from about 3 million to nearly 12, and those between 45 and 64 years of age from nearly 10½ to 31 million (**6**). However, age discriminations in employment limit the potential productivity of many older individuals. Increases in mandatory pension plans may operate to reduce still further the work opportunities available to older persons. Parran estimates that 1,500,000 persons are prematurely retired and that as a result $4,500,000,000 a year is lost in productive earning capacity (**3**). Life has been made longer, without provision for extending productivity. However, these issues are not being ignored. The National Conference on Ageing identified numerous problems of oldsters (**1**), the Desmond committee of the New York Legislature has stressed especially employment problems in the older years (**2**), a recent issue of the *Annals of the American Academy of Political and Social Science* deals with oldsters in the community and in industry (**5**), and continuing studies by the United States Employment Service concern the employability of older persons. These and other reports emphasize the undesirability of prevailing practices regarding hiring and retirement in relation to age, and suggest various means by which even a person somewhat handicapped by age may be kept on the job, as by transfer to less exacting duties, part-time employment, or establishment of departments specifically for older workers (**4**, **7**).

A major research need is clearly for direct and practical appraisals of the competence of older as compared to younger workers in a variety of actual working situations. Such findings could become basic in the formulation of retirement policies, job adjustment to facilitate continued employment of older workers, selection of older persons for employment, and induction and training plans for older workers. In the absence of such findings, administrative actions affecting older workers will continue to be based on common impressions, half-truths, and special cases. The present investigation attempts a simple appraisal of the comparative performance of industrial workers of different age groups in a large organization, and a rough analysis of factors involved.

Cases and Materials

This study is concerned with 3,162 workers of both sexes, ranging in age from 18 through 76 and performing various duties in an organization that includes a variety of operations. And the workers covered ranged from foremen and minor executives (more important executives and supervisory personnel were not included in the cases studied), skilled craftsmen, operators of both heavy and light machinery and equipment, and inspectors to clerks, and to unskilled labor. In total it is believed that the variety of work and types of workers covered were as great as would be found in most plants with payrolls of four or five thousand employees. The supervisory methods and records were judged rather better than in many industrial plants. The cooperation of the management was excellent. All data were obtained from the personnel records maintained for each employee, which included age, date of hiring, and appraisals by foremen concerning competence. All this material was recorded in brief by the investigators on summary cards. Appraisals were made when administrative actions were taken affecting employees, upon completion of training courses, as a basis in recognition for outstanding service, in connection with a grievance of an employee, and more generally where some record seemed appropriate. For many employees, several appraisals were found, made by various supervisors. All of these were considered.

* The writer was assisted in this investigation by Mark Weldon Smith, Research Fellow, The Ohio State University Development Fund. The project was under the direction of Dr. S. L. Pressey.

The appraisals consisted of highly informal statements made by the foremen and covered, for one worker or another, approximately 300 terms covering a wide range of abilities, characteristics and traits. It is felt that the evaluations were of especial descriptive worth, since they represented unstructured statements made by supervisors, in their own language, with reference to each worker as he was observed in his working situation. The purpose of these records was not primarily to determine promotions or other administrative actions which so obviously and promptly affected the worker as to tend to make the ratings biased, but rather to provide meaningful appraisals for long-time policy, which could be referred to by the personnel department on occasions when a descriptive picture of the particular worker was desired.

To reduce the number of terms in these appraisals to a workable list, those which were ambiguous or infrequently mentioned were eliminated or combined with others of similar meaning, after consultations with personnel representatives and foremen and careful consideration by the writer and several associates. The 300 terms were finally reduced to a basic list of eight abilities, eight character traits, and four common faults which were felt to be largely unambiguous and fairly specific in the working situation. In handling the data, the first step was to find the numbers of persons in each age group (and of each sex) who were reported by foremen to have any of the twenty traits. Minus values were subtracted; thus, the number called inefficient was subtracted from the number called efficient. Then, net figures were converted into net percentages.

Results

Table 1 shows net percentages of men and women in each age group with over two years' service—and thus presumably in the organization long enough to have been farily well appraised—who had been mentioned by foremen as having each trait. Age groupings are in accordance with common census practice.

Table 1

Percentages of Employees with Over Two Years' Service Mentioned by Foremen as Having Specified Traits

	Ages of Men				Ages of Women			
	18–29	30–44	45–59	60 up	18–29	30–44	45–59	60 up
Ability								
Efficiency*	49	41	40	40	41	37	41	48
Job Knowledge	15	13	12	11	8	12	12	8
Ability to Learn*	12	7	2	3	17	6	11	4
Accuracy	8	7	4	3	34	15	22	20
Initiative	6	7	5	5	11	5	16	4
Character								
Cooperativeness*	32	28	25	24	24	19	29	24
Dependability	27	26	22	26	26	19	30	24
Attendance*	5	10	13	15	3	5	12	24
Thoroughness	5	3	3	4	6	5	10	—
Steadiness*	10	14	19	19	8	13	27	28
Conscientiousness*	10	8	10	14	8	6	17	20
Tactfulness	15	11	11	6	16	14	12	20
Faults								
Slowness*	−1	−2	3	12	−9	−5	2	4
Instability	—	2	3	2	4	6	12	4
Total	280	644	538	316	115	242	165	25

* Traits showing a difference of 5 or more between the first two as compared to the last two age groups. The following items were omitted as showing no net percentage in any age group for either sex of 10 or more: job satisfaction, promotion possibilities, supervisory ability, work without supervision, physical difficulties and overtalkativeness. Thus, the highest mention of physical difficulties was 6 for men 60 and over.

The data would seem to indicate that age differences in traits were relatively small. Older workers of both sexes with over two years' service were reported to learn less readily and to be slower, but also more frequently to show good attendance, steadiness, and conscientiousness. These trends are quite consistent; thus women under 30 were fairly often mentioned as rapid workers (giving a net of −9 per cent slow), but the per cents drop steadily to 4 per cent slow over 60. In the other traits covered, age differences appear to be negligible, or not consistent. For example, the difference in efficiency between the oldest and youngest group of men is 9 net percentage points favoring the younger, but there is only a 1 point drop between the 30–44 and 60 up group. However, the oldest women are considered more efficient than the younger by 7 points. No consistent age differences appear in job knowledge, accuracy, dependability, emotional stability. As mentioned in the footnote to Table 1, physical difficulties were not mentioned often enough even for the oldest group (6 per cent mentions for men over 60) to indicate that such handicaps were of importance.

The preceding table deals with traits, not persons. For practical purposes it seemed desirable so to combine information on each person that a total appraisal of his competence could be obtained. A net evaluation of each worker was therefore obtained by summing favorable characterizations and subtracting the unfavorable ones in terms of the above mentioned traits, with special weighting for the

Table 2

Distributions of Net Appraisals of Competence for Men and Women with Over Two Years' Service, and Men with Under Two Years'*

Net Appraisal	Men Over 2 Years				Women Over 2 Years				Men Under 2 Years			
	18–29	30–44	45–59	60 up	18–29	30–44	45–59	60 up	18–29	30–44	45–59	60 up
13		1					1					
12		1	1		5		1					
11		4	2		1	1	1			1		
10	1	5	2		1	2	2				2	
9	4	12	10	5		7	3	1			3	1
8	9	11	7	5	5	8	5	3	1	3		
7	10	27	24	9	4	5	10	2	5	5	2	
6	18	48	32	25	10	13	13		7	12	8	1
5	30	44	26	20	9	17	12	1	25	28	14	5
4	37	60	40	22	**8**	21	13	**2**	29	40	15	4
3	**28**	**69**	**73**	21	16	**31**	**18**	3	30	40	14	**6**
2	20	51	39	**27**	5	15	10		32	**36**	**15**	2
1	18	52	31	27	8	9	17	3	**32**	22	13	1
0	12	23	32	20	2	11	5	1	27	21	9	5
−1	18	24	13	19	4	10	9		17	20	7	
−2	6	14	7	6	1	8	2		16	7	3	1
−3	5	12	5	6		3	4		3	6		
−4	1	3	2	1	3	1	1	1	2	3	1	
−5		3	1				1		2	1	1	
−6				1			1					
−7												
−8		1										
Total	217	465	347	214	82	162	129	17	228	245	107	26
Mean	3.2	3.2	3.2	2.7	4.1	3.3	3.4	4.0	1.9	2.4	2.8	3.0
% Over 3	50	47	42	40	52	46	47	53	30	36	41	42
% Under 3	0	2	1	1	4	1	2	6	2	2	2	0

* Figures in boldface indicate intervals in which the means fall.

important factor of efficiency (3 points were given if efficiency was stated to be outstanding, 2 if very good, 1 if good or average, −1 if poor, −2 if unsatisfactory, and −3 if dismissed or disciplinary action had been taken because of inefficiency). Table 2 shows distributions and means of these net evaluations in the various age groups. Data are shown again for those with over two years' service and also for men with less time in the organization; women with less than two years' service were omitted because few of them were in the older groups, and many had not been appraised.

The wide range of scores for both men and women suggests that the evaluative procedure did discriminate between good and poor employees. The means show no large differences between age groups. Inspection of the total distributions also indicates little difference in the extremes. The comparatively small number so good as to obtain favorable comments totaling 10 or more are all under 60. But 40 per cent of the men 60 and over with two years' service scored over 3, as compared with 50 per cent for the youngest group. Women with over two years' service showed 52 per cent in the 18–29 group so scoring as compared with 53 per cent of the few cases 60 and over.[1] None of the age groups showed any appreciable number of individuals scoring less than minus three. Both the means and the distributions for the various age groups, of the net evaluations for workers with over two years' service, seem to indicate the value of retaining older workers.

In Table 2 the last four columns deal with men hired during the past two years. These men show both a steady rise in means with age (from 1.9 to 3.0) and a larger percentage of older workers scoring over 3. These findings for those recently hired are felt to be very important. If workers hired even in their sixties are found satisfactory, it would appear that a large labor pool of older workers could be turned to in time of need.

Table 3 gives distributions and means of net appraisals of employees 60 and over, in a breakdown by five-year intervals. Cases over and under two years' service are here combined.

Table 3

Distributions of Net Appraisals of Competence of All Employees 60 and Over for Those Both Over and Under Two Years' Service*

Net Appraisal	Males			Females†	
	60–64	65–69	70-up	60–64	65–69
9	4	2			1
8		4	1	3	
7	6	3		2	1
6	19	6	1		
5	19	3	3		1
4	19	3	4	**3**	**1**
3	20	**6**	**1**	3	
2	**20**	8	1	1	
1	20	5	3	3	
0	13	9	3		1
−1	15	4			1
−2	7				
−3	5	1			
−4	1				1
−5					
−6	1				
Total	169	54	17	15	7
Mean	2.5	3.1	3.1	4.3	2.9

* Figures in boldface indicate intervals in which the means fall, with the exception of the last column in which the interval containing the median is boldfaced.
† No women were listed as 70-up.

A considerable number of men over 65, and some even 70 and over, were on the payroll; and none in this oldest group was given a minus appraisal. A large proportion of all those of both sexes, 65 and over, appear to have been distinctly satisfactory.

Major Conclusions

The organization in which this study was made employs many older workers, including an appreciable number over 65. That most of these older workers were considered competent on the job and compared favorably with younger workers in the organization would seem a justification for their employment, evidence of wise personnel policy in their use, and evidence against arbitrary retirement at 65.

That men hired for the first time when over 45 and even when 60 or over (male workers

[1] Because of the confidential nature of the personnel records, evaluations of key supervisors were not made. Many of these supervisors are older persons; had evaluations of them been included, the net percentages favoring the older age groups might have been higher.

with less than two years' service in the upper two age groups) were given higher net appraisals than younger men currently being hired seems even more striking vindication of the older worker as being competent. The common business practice of refusing to hire a new worker over 45 seems even contra-indicated—at least for some types of work, since the older applicant seems to be the better risk! At all ages there are great individual differences. Clearly, selection should be on other bases than mere chronological age.

Certain changes in traits with age, and large individual differences in traits at every age, all indicate the need for individualized selection and placement, and for job readjustment of some with increasing age or because original placement was in some respect unsatisfactory. Many such readjustments were made in this organization. More such readjustments might be expected to make older workers even more satisfactory.

Summary

Current census data suggest that by 1960 there will be nearly 15 million persons aged 65 and over, and 35 million between ages 45 and 64 or a total of 50 million over age 45. From the standpoint of economic balance and defense, the manpower potential of an increasing older population should be carefully studied.

The present investigation involved a detailed appraisal of the personnel records of several thousand workers between the ages of 18 and 76 in a large organization. The frequency of mention of each common trait at each age was determined, and the net total of favorable and unfavorable traits mentioned by supervisors for each employee was used as an index of his worth. The following findings stood out:

1. Older workers were mentioned somewhat more frequently than younger as not learning readily and as being slow; but the older workers were considered to be better in attendance, steadiness, and conscientiousness. The remaining traits showed no consistent changes with age.

2. Net appraisals of workers showed negligible differences between age groups but great individual differences at every age. Employees over 65, and even 70 and over, were appraised very favorably. It would seem clear that arbitrary retirement at 65 or hiring discriminations against employees because of age have little warrant.

3. The organization has in the last two years taken on many workers over 45 and even over 60. Most of these were appraised favorably. Age limits in hiring seem clearly questionable.

Quite evidently workers should be employed and retained on the basis of merit without reference to age. Biases and misconceptions limiting the use of older persons should be replaced by facts. Oldsters can maintain productivity, thus making an extended productive life worth while, strengthening manpower resources, and lessening possible economic burdens resulting from dependency of large numbers of non-productive older persons.

Received July 7, 1952.
Early publication.

References

1. National Conference on Ageing. *Man and his years.* Health Publications Institute, Raleigh, North Carolina, 1951.
2. New York Joint Legislative Committee on Problems of the Ageing, Albany, N. Y. *Never too old,* Legislative Document No. 32, 1949; *No time to grow old,* Legislative Document No. 12, 1951.
3. Parran, T. Must you retire at 65? *Colliers,* 1952, **129**, 15.
4. Stanton, Jeannette E. Part-time employment for the older worker. *J. appl. Psychol.,* 1951, **35**, 418–421.
5. Tibbets, C. (Ed.) Social contributions by the ageing. *Ann. Amer. Acad. pol. soc. Sci.,* 1952, **258**, 279.
6. U. S. Department of Labor, Bureau of Labor Statistics. *Fact book on the employment problems of older workers.* Prepared for the Conference on Ageing, Washington, D. C., August 13–15, 1950.
7. Welford, A. T., and Speakman, D. *The employability of older people. The aged and society.* Champaign, Illinois: Twin City Printing Company, 1950, 198–199.

Relation between *How Supervise?*, Intelligence and Education for a Group of Supervisory Candidates in Industry*

Frederic R. Wickert

Michigan State College

Informal impressions indicate that the File and Remmers questionnaire *How Supervise?* (3) continues to be rather widely used to measure the attitudes and knowledges required for supervisory success, notwithstanding its apparent deficiencies as a measuring device for this purpose. One rather obvious difficulty likely to be encountered in interpreting the test scores of industrial supervisors is, of course, that the scores may be measuring facility with words, or educational level, or verbal intelligence, rather than supervisory abilities or aptitudes.

The literature concerning the relationship between *How Supervise?* and educational level and intelligence at first glance has not, until recently, been too informative.

File (1) reported a correlation of +.35 (N = 577) between *How Supervise?* scores and education. He said with respect to this correlation, "The optimum amount of correlation which should exist between education and a test of supervisory quality is problematical. While such a test should not correlate highly with amount of education, doubtless, formal education does provide valuable learning situations which are generally helpful." He further indicated that much of the positive relationship he found between scores on his test and education was due to the difference in test performance between those with college as compared to those without college. File explained this relationship without any reference to the possible influence of verbal intelligence or of training in thinking through verbal problems or of selection of persons with verbal skills. He said, ". . . Selection at the college level tends to 'weed out' individuals who have failed to develop an understanding of the general factors of human relations, or that colleges . . . provide considerable opportunity for gaining insight into human relations problems."

Sartain (6) reported a correlation of −.44 between the *Adaptability Test* and an "experimental" form of *How Supervise?* which was scored in such a way that "low scores" indicated "a favorable attitude." The N was 40. In the text of his article, Sartain stated that his results indicated "that general mental ability goes with favorable supervisory attitudes to a moderate degree." Sartain's negative correlation coefficient evidently indicated a positive relationship.

File (2), in an article following Sartain's report, attempted to explain the difference between his finding of a positive correlation (+ .35) between *How Supervise?* and education and Sartain's negative correlation (−.44) between the *Adaptability Test* and *How Supervise?*, without apparently appreciating that Sartain's negative correlation meant that a positive relationship really existed. Accordingly, Sartain's findings do not contradict those of File, and File need not have been so concerned about reconciling the results of the two studies.

Two recent articles have done much to clarify the interrelationships with which this paper is concerned. In the first article, Millard (5) reported that for office (higher level) supervisors the correlation between scores on *How Supervise?* and the *Adaptability Test* was + .22. However, the correlation for a group of factory foremen (lower level or first line) was + .71.

The second recent article is that of Maloney (4). He made a Flesch analysis of *How Supervise?*. His findings were: "The mean readability for both forms, directions included, is of high school graduate difficulty." Maloney then suggested that Millard's high correlation between *How Supervise?* scores and scores on the *Adaptability Test* for foremen was not

* Dr. Glen Grimsley and Dr. John F. MacNaughton participated principally in those aspects of this project which had a bearing on giving service to the client company.

surprising since both *How Supervise?* and the *Adaptability Test* measure the ability to understand written material. Maloney concluded, ". . . for lower level personnel *How Supervise?* is of doubtful validity as a measure of supervisory ability."

The present study most resembles that of Millard. However, rather than reporting correlations between *How Supervise?* and an intelligence test for two groups obviously differing not only in educational level but also in social status, type of work supervised, etc., the present study reports the relationship between *How Supervise?* and education and intelligence for but one group of subjects and then divides the total group into upper and lower educational groups.

The data here reported are only a small part of the data collected in connection with a larger study. This larger study involved, in part, administering a number of psychological tests to (including Form A of *How Supervise?* and the Advanced Short Form of the California Mental Maturity Test), as well as obtaining certain biographical data on, a group of almost 100 candidates for supervisory positions. These men were employed, almost all in the shop rather than the office, of one plant of a large metal products company.

Results

Table 1 shows the interrelationships among scores on *How Supervise?*, the "language factors" score on the Advanced Short Form of the California Mental Maturity Test, and years of education. All three of these variables appear to be rather substantially related to each other for the total group of supervisory candidates tested in this study.

Table 1

Intercorrelations among *How Supervise?* Scores, Scores on the Language Aspects of a General Intelligence Test, and Years of Education

Tests and Education	Variables 1	2	3
1. *How Supervise?*		+.66*	+.54
2. Verbal Intelligence	+.66*		+.59
3. Years of Education	+.54	+.59	

* N = 95. All other correlation coefficients are based on an N = 87. (All correlations are tetrachoric because these data were lifted from a large table of intercorrelations where the numbers of coefficients to be computed precluded the use of more refined measures of correlation.)

Table 2

Relation between *How Supervise?* and Verbal Intelligence for the Well Educated and the Less Well Educated Supervisory Candidates

Educational Level	Tetrachoric *r* between *How Supervise?* and a Verbal Intelligence Test	N
12th grade education or more	+.20	57
11th grade education or less	+.65	30

In order to analyze these interrelationships further, Table 2 was prepared.

Table 2 shows the relationship between *How Supervise?* and verbal intelligence with the effect of education crudely partialed out. It is clear from Table 2 that among the better educated group there is comparatively little relationship between *How Supervise?* scores and intelligence. A high degree of supervisory knowledge of the sort measured by *How Supervise?* does not seem to depend particularly on good verbal intelligence among this better educated group.

For the less well educated group, however, *How Supervise?* scores and intelligence test scores are quite closely related. Ability to get a high score on *How Supervise?* among this group would appear to depend to an appreciable extent on possessing good verbal intelligence.

Discussion

The data reported here are in striking agreement with those of Millard. In addition, they fit in nicely with Maloney's findings. It will be recalled that Maloney reported that a Flesch analysis put the readability level of *How Supervise?* at the high school graduate level; the present study found the best cutting point between those subjects whose *How Supervise?* scores were highly related to intelligence test scores and those subjects for whom this relationship was negligible, was also at just about the high school graduate level.

The finding of the present study that *How*

Supervise? scores and verbal intelligence are rather closely related to each other among the subjects with less education suggests strongly that *How Supervise?* scores for these people should be interpreted with caution. On the surface, *How Supervise?* might almost as well provide a measure of general verbal intelligence as a measure of supervisory knowledge and understanding among this group with relatively little education. Maloney's findings, however, point to the interpretation that for this low education group the high relationship between *How Supervise?* and intelligence test scores exists not because *How Supervise?* is an intelligence test, but because, for the less well educated persons, readability is the important factor in taking *How Supervise?*. Maloney's statement was, "To the extent to which educational attainment is an indication of reading ability the test can be read only with great difficulty by the average factory foreman. Therefore, for lower level personnel *How Supervise?* is of doubtful validity as a measure of supervisory ability."

The further finding that *How Supervise?* scores and verbal intelligence are only slightly related, when the test scores are those for subjects with more education, suggests that scores on the present forms of *How Supervise?* can be used with some confidence that they may be measuring supervisory knowledge and understanding apart from general verbal intelligence (and readability) when one is working with persons who went through at least the twelfth grade.

Summary

This study summarizes past data and presents further data on the interrelationships between *How Supervise?* (a widely used test of supervisory knowledge), verbal intelligence, and amount of education. The subjects for this study were almost 100 candidates for shop supervisory positions in a metal products company.

Results showed that these three variables were more highly interrelated than early studies had reported. Further analysis of the results of the present study corroborated a rather recent study by Millard in that *How Supervise?* and verbal intelligence test scores were quite highly related among the less well educated of the total group tested. In fact, the relationship was so close that there seemed to be a real question with respect to whether *How Supervise?* could measure supervisory knowledge apart from general intelligence for this less well educated group. Another recent study by Maloney, however, suggested that results like Millard's (and inferentially those of the present study) could be accounted for by the fact that *How Supervise?* is difficult to read for those whose educational level is relatively low.

The general conclusion is that *How Supervise?*, for those persons who did not graduate from high school, measures intelligence (or readability of *How Supervise?*) rather than knowledge of the principles of supervision. However, for relatively well educated persons, *How Supervise?* scores have little relationship with intelligence test scores (or readability). When testing this better educated group, there would seem to be some assurance that *How Supervise?* can measure supervisory knowledge apart from verbal intelligence (or the readability of *How Supervise?*).

Received November 19, 1951.

References

1. File, Q. W. The measurement of supervisory quality in industry. *J. appl. Psychol.*, 1945, **29**, 323–337.
2. File, Q. W., and Remmers, H. H. Studies in supervisory evaluation. *J. appl. Psychol.*, 1946, **30**, 421–425.
3. File, Q. W., and Remmers, H. H. *How Supervise?* (revised manual). New York: Psychological Corporation, 1948, pp. 8.
4. Maloney, P. W. Reading ease scores for File's *How Supervise?*. *J. appl. Psychol.*, 1952, **36**, 225–227.
5. Millard, K. A. Is *How Supervise?* an intelligence test. *J. appl. Psychol.*, 1952, **36**, 221–224.
6. Sartain, A. Q. Relation between scores on certain standard tests and supervisory success in an aircraft factory. *J. appl. Psychol.*, 1946, **30**, 328–339.

Changes Occurring in Teacher-Pupil Attitudes during a Two-Weeks Guidance Workshop

Jack Shaw, Herbert J. Klausmeier, Arno H. Luker, and Howard T. Reid

Department of Educational Psychology, Colorado State College of Education

Several investigations in recent years using the *Minnesota Teacher Attitudes Inventory* have shown that teacher-pupil relationships are closely related to teacher-pupil attitudes, that teacher-pupil attitudes can be validly measured, and that the effects of teacher education and teaching experience on these attitudes can be determined (**1**, **3**, **4**, **5**, **7**). The reliability of the MTAI (Form A) was reported to be .82 based on test-retest data one week apart (**2**). The validity of the MTAI was determined empirically with a group of experienced teachers by its authors who found that it correlated +.60 with a criterion of the combined ratings of principals, ratings of observations by one of the authors, and ratings by the pupils of the teachers. Further evidence of validity of the inventory was found by Shaw using a criterion of practice teaching supervisor's ratings (**9**).

The research by the authors of the MTAI indicated that during the pre-service education of teachers statistically significant changes in teacher-pupil attitudes did occur. Nothing has appeared as yet in the literature, however, to indicate whether or not changes in teacher attitudes occur during post-graduate training of experienced teachers. The authors of this article undertook to investigate this problem among experienced teachers, counselors, and administrators enrolled in a two-weeks guidance workshop.

Subjects

The subjects of the study were 158 teachers, counselors, principals, and superintendents enrolled in a two-weeks guidance workshop held at Colorado State College of Education in June, 1951. The workshop was taken for three quarter-hours' credit towards the B.A., M.A., or Ed.D. degrees. These students came from 23 states and Hawaii. There were 99 men and 59 women in the class. One student had freshman standing, 4 were juniors, 9 were seniors, 110 held bachelor's degrees, 33 held master's degrees, and one had earned a specialist's diploma one year beyond a master's degree.

Their experience in the educational profession (including teaching, counseling, supervision, and administration) ranged from 0 to 35 years, with a mean of 9.8 years and a median of 7 years. Only 12 students had no teaching experience in the field. The hours of course work in educational psychology as reported by the members of the workshop ranged from 0 to 105 quarter hours with a median of 15 quarter hours. Their major fields of study were reported as follows: educational psychology, 23 members; educational administration, 30; elementary education, 49; and secondary education, 56.

Some indication that the teacher-pupil attitudes of these subjects at the time of registration in the workshop were not widely atypical of graduate students in education is seen by comparison of the group with a norm group of 115 graduate students in education described in the manual of the MTAI (**2**). The workshop group, consisting of 158 members, had a mean score of 53.0 on the MTAI with a standard deviation of 25.5. The norm group reported in the MTAI manual, consisting of 115 members, had a mean score of 48.2 with a standard deviation of 27.5.

The Problem

Specifically, the problem under study was twofold: 1. What are the differences in teacher-pupil attitudes as measured by MTAI among the teachers enrolled in the guidance workshop in relation to their major fields of study? 2. What changes in scores on the MTAI occurred during the two-weeks workshop (a) among the group as a whole, and (b) among teachers grouped according to major fields of study?

Procedures

The MTAI (Form A) was administered to the teachers in the workshop at the beginning of the two-week period, prior to any lecture, discussion, or work on the workshop topic, "The Improvement of Guidance Services in the Schools of America." The workshop members were told that the inventory was being administered as a good teaching procedure whereby the faculty of the workshop might better understand the teachers with whom they were going to work for two weeks. Scores (in decile ranks based on MTAI norms provided by the inventory authors for graduate students in education) were posted by student initials on the bulletin board. Without advance notice, the MTAI (Form A) was again administered on the last day of the workshop. The students were assured its purpose was to evaluate the proceedings of the workshop and in no way would the scores of the inventory be used in assigning marks for course credit.

Results

The results of the initial and final administrations of the MTAI are shown in Table I. Examination of the table shows that the four subgroups and the total group made statistically significant increases as measured by the difference in the means of the initial and final inventory scores. The total group (N is 158) showed changes in attitudes in a favorable direction from a mean of 53.0 to 69.2, the difference 16.2 being statistically significant at better than the one-tenth of one per cent level.[1]

The order of greatest mean gain by major groups was: secondary education, 18.7; educational administration, 18.4; educational psychology, 14.6; and elementary education, 12.5.

The order of the groups according to means on the initial administration was: educational psychology, 65.1; elementary education, 57.6; educational administration, 49.3; and secondary education, 46.1. The same order was found in the second administration of the inventory, although the differences between the means of the successive groups were reduced except between educational psychology and elementary education.

The authors of the inventory reported that changes in attitude scores accompanied professional education courses in the pre-service education of students for the teaching profession (**1, 4**). The results of this study indicate that the attitudes of experienced teachers returning for refresher or graduate training changed in a favorable direction during a two-weeks guidance workshop.

Implications

The results of this study further support the findings of the authors of MTAI that professional courses in education can result in change in attitude scores. Whereas previous findings indicated that such changes seemed to occur primarily in the early professional

[1] Test of significance of difference between means for correlated data was used.

Table 1

Comparison of the Scores Made by the Workshop Members on the Minnesota Teacher Attitudes Inventory on the Initial Test and Retest according to Major Field of Study

	Major Field of Study									
	Ed. Psych.		Ed. Admin.		Elem. Ed.		Sec. Ed.		Total Group	
	Initial Test	Re-test	Initial Test	Re-test	Initial Test	Re-test	Initial Test	Re-test	Initial Test	Re-test
N	23	23	30	30	49	49	56	56	158	158
Mean	65.1	79.7	49.3	67.7	57.6	70.1	46.1	64.8	53.0	69.2
S.D.	27.9	11.8	22.4	20.0	17.7	18.9	28.2	22.0	25.5	16.0
$t_{(M_I - M_R)}$	2.66		4.40		3.43		6.13		11.17	
Probability less than	.02		.001		.01		.001		.001	

courses in a teacher-education program, this study indicated that changes occurred in post-graduate courses in education.

Furthermore the results of this study support the findings of the authors of the inventory that the MTAI differentiates among teachers of different grade level. Whereas it was previously shown that the order of the better attitudes among teachers was first early childhood teachers, second intermediate grade teachers, third senior high school teachers, and fourth junior high school teachers, this study showed a hierarchy according to field of study as follows: (1) educational psychology; (2) elementary education; (3) educational administration; and (4) secondary education. Possibly these orders represent some form of a continuum from interest in children to interest in academic subject matter. This continuum in one study was found to exist among elementary and secondary teaching candidates in the sophomore year of their college education (**6**).

Because a previous investigation (**1**) found a deterioration in attitudes as measured by MTAI among teachers during their first year of teaching experience, the question arises whether or not the gains in attitude scores made during professional courses reflect real changes in attitudes or merely an increase in facility with attitudinal terminology which may readily be forgotten after a few months away from the educational institution. The present study should therefore be followed up at the end of the subsequent school year to determine whether or not the gains in favorable attitudes of experienced teachers made in a two-weeks workshop are lost during the following year of additional experience in the field.

Finally it has been demonstrated by this study that the MTAI may be a useful instrument in evaluating the effectiveness of procedures or techniques such as a workshop in the professional education of teachers, counselors, and administrators.

Received October 10, 1951.

References

1. Callis, R. *Change in teacher-pupil attitudes related to training and experience.* Unpublished Ph.D. thesis, University of Minnesota, 1948.
2. Cook, W. W., Leeds, C. H., and Callis, R. *Manual for the Minnesota teacher attitude inventory, form A.* New York: The Psychological Corporation, 1950.
3. Cook, W., and Leeds, C. H. Measuring the teaching personality. *Educ. psychol. Measmt,* 1947, **7**, 399–410.
4. Cook, W. W., Leeds, C. H., and Callis, R. Predicting teacher-pupil relations. *The evaluation of student teaching.* 28th Yearbook, The Association for Student Teaching, 1949, Ch. 4.
5. Fuller, Elizabeth M. The use of teacher-pupil attitudes, self-rating and measures of general ability in the preservice selection of nursery school-kindergarten-primary teachers. *J. educ. Res.,* 1951, **14**, 675–686.
6. Klausmeier, H., Luker, A., and Stromswold, S. Factors influencing choice of teaching career among college sophomores. *J. educ. Res.,* 1951, **45**, 23–32.
7. Leeds, C. H. A scale for measuring teacher-pupil attitudes and teacher-pupil rapport. *Psychol. Monogr.,* 1950, **64**, No. 6 (Whole No. 312).
8. Leeds, C. H., and Cook, W. W. The construction and differential value of a scale for determining teacher-pupil attitudes. *J. exp. Educ.,* 1947, **16**, 149–159.
9. Shaw, J. The function of the interview in determining fitness for teacher-training. *J. educ. Res.,* 1952, **45**, 667–681.

Personality Characteristics Related to Success in Practice Teaching

Harrison G. Gough

University of California, Berkeley

and

William H. Pemberton

Consulting Psychologist, San Francisco, California

The importance of personality characteristics for tasks involving personal interaction, leadership, and social understanding is incontestable. The difficulty in utilizing a principle such as this lies more in devising techniques and methods for its adequate application than in proving the truth of the basic assumption. Advances in the methodology of personality assessment and evaluation have yielded various instruments which show promise of overcoming this technological barrier. This study is concerned with the application of one of these devices, the Minnesota Multiphasic Personality Inventory (**3**), to the problem of predicting success in student practice teaching.

This investigation is a revision and extension of an earlier one carried out by Pemberton (**4**) as a doctoral dissertation in the School of Education at the University of California, Berkeley. In that project a sample of 96 males enrolled in a course in secondary school practice teaching was studied by means of the MMPI and the group Rorschach. Neither scores on MMPI scales nor conventional scoring categories on the Rorschach yielded much in the way of significant relationships to practice teaching ratings.[1]

At the same time, a clinical-intuitive sorting of the MMPI profiles (by H. G. G.) did reveal a positive correlation with the ratings. The results of this sorting are presented in Table 1.

The favorable outcome for this procedure suggested that the MMPI profiles did, in fact, depict significant facets of personality, but that the relationships were obscured by a conventional scale-by-scale analysis. This inference is in line with an abundance of previous work with the MMPI[2] which has underscored the necessity of considering patterns and configurations of scores in interpreting test profiles. The results of the present paper, it might be emphasized, lend further support to this admonition.

Table 1

Evaluation of Clinical Sorting of MMPI Profiles of Male Student Practice Teachers

		Practice Teaching Ratings			
		Higher Ratings	Average Ratings	Lower Ratings	
Sortings	Higher	9	14	3	26
	Average	13	20	11	44
	Lower	4	10	12	26
		26	44	26	96

$\chi^2 = 8.51$; d.f. = 4; and $P > .05, < .10$.

In the earlier study the test results of the female students were not considered. They were later scored, and the profiles given to another clinical psychologist[3] for ratings on "potential success in practice teaching." The correlation of the clinical ratings with the criterion ratings for the 58 profiles involved was +.24, S.E. .13. This prediction was achieved despite the fact that no single scale on the MMPI showed a significant relationship to the criterion ratings.

The decision was made, accordingly, to attempt to submit the intuitive cues used in test interpretation to analytical and quanti-

[1] Students in practice teaching were rated twice during the semester on each of four variables: (1) Personal relations with students and teachers; (2) Command and use of subject matter; (3) Teaching skill; (4) Class management. The ratings were expressed in qualitative terms: outstanding, excellent, good, fair, and poor. In Pemberton's study only ratings on variables (1) and (4) were used, in an effort to maximize the salience of personality characteristics. In the present investigation the qualitative ratings were transformed into a scale of 5, 4, 3, 2, and 1, and the entries were then summated over the eight (4 × 2) ratings. The total scores thus derived served as criteria for the study.

[2] See for example: Gough, H. G. Diagnostic patterns on the MMPI. *J. clin. Psychol.*, 1946, **2**, 23–37.

[3] The writers are indebted to Dr. George S. Welsh for furnishing these ratings.

tative investigation to determine whether any objective signs or indices on the MMPI could be isolated which would possess useful predictive power for the criterion of success in practice teaching.

The first step was to categorize both the male and female samples into those with higher, average, and lower ratings, and then to calculate mean scores for each of the MMPI scales available. Some of the original cases had not taken the complete form of the MMPI, and as we desired to use a number of nonpathological scales[4] these records were eliminated. The total sample included 89 males and 58 females.

Table 2 presents the results of this analysis for the female sample.[5] None of the 22 MMPI scales considered yielded a significant F-ratio when evaluated in this fashion.

The results for the male sample appear in Table 3. Three of the scales show some promise in this analysis. The more successful male practice teachers tend to score lower on Hy (hysteria), lower on Pd (psychopatic deviate), and higher on Psy (psychological aptitude). Besides the scales listed in the two tables, two of the indices recently proposed by Welsh (5) were evaluated. Neither the anxiety index (AI) nor the internalization ratio (IR) revealed a significant relationship.

The results with individual scales confirmed the findings of the earlier analysis and indicated that any relevant factors on the MMPI would have to be identified by a more configural and intuitive approach. The steps involved in translating this deduction into a research design were: first, a great deal of free association and ratiocination about the signs and clues involved in the intuitive sortings; and second, an attempt to write simple indices and functions summarizing them. Altogether some 15 "signs" were devised and tested against the threefold criterion breakdown.[6] Eight of these signs revealed some dicriminatory potential. These signs are listed in Table 4.

Table 4

MMPI "Signs" Predictive of Success in Practice Teaching*

Sign	χ^2	d.f.	P
1. Pa≥50, ≤56	7.33	2	>.02, <.05
2. Ma≥48, ≤60	6.74	2	>.02, <.05
3. Pa+L≤56	3.27	1	>.05, <.10
4. Ma≥48, ≤60 *and* Pd≥46, ≤58	2.88	1	>.05, <.10
5. Hy≥Pd+1†	1.69	1	>.10, <.20
6. K+Ie+3 Do≥ Pt+Sc+Pa+15†	1.22	1	>.20, <.30
7. St+Re+Ie≥102†	2.84	1	>.05, <.10
8. Ma+K+D≥ Pd+Hs+Pt	1.75	1	>.10, <.20

* All scales are expressed in T-scores, except L, Ie (intellectual efficiency), St (status), Do (dominance), and Re (responsibility).

† These constants were selected so as to divide the sample of 147 as evenly as possible into those with sign present and those with sign absent.

After these eight signs had been established each of the original records was scored for presence or absence of each sign, with a total "sign-score" also being obtained. With eight signs, this score could range from zero to eight. As would be expected, these "sign-scores" were significantly related to success in practice teaching. The results are given in Table 5.

A separate analysis was made for each sex. For females the χ^2 was 7.3, d.f. 4, $P > .10$, $< .20$. For males the χ^2 was 9.3, d.f. 4, and $P > .05$, $< .10$. For the combined sample only 6 of the 147 cases are misclassified by two steps, and 71 cases are exactly classified on the diagonal. The over-all predictive efficiency would thus appear to be good for this sample.

A cross validation of these signs was also attempted. MMPI profiles of students enrolled in courses in primary grades practice teaching were obtained.[7] Eighty cases (40 of each sex) with quite high ratings were selected, and a second sample of 80 (40 of each sex) with rather low ratings was also chosen. These students had not taken the full form of the

[4] The additional scales used were: Re (social responsibility), Do (dominance), St (status), Sp (social participativeness), Pr (prejudice), Ac (academic achievement, high school), Ds (dissimulation), Psy (psychological aptitude), Ie (intellectual efficiency), and Si (social introversion). These scales have been described in a series of publications too long to be enumerated here. As examples the reader is referred to **1** and **2**.

[5] To reduce printing costs Tables 2 and 3 have been deposited with the American Documentation Institute. Order Document 3661 from American Documentation Institute, 1719 N Street N.W., Washington 6, D. C., remitting $1.00 for microfilm (images 1 inch high on standard 35-mm. motion picture film) or $1.00 for photocopies (6 × 8 inches) readable without optical aid.

[6] The two subsamples were pooled for this analysis, providing a total of 147 cases.

[7] Dr. Fred Tyler very kindly made these records available.

MMPI, so we were unable to compute the three signs involving K, Ie, Do, St, and Re. The results of this analysis are presented in Table 6.

The χ^2 here is significant between the .02 and .05 levels. From Table 6 it appears that the signs are more effective in identifying better practice teachers than in specifying poorer ones. In fact, there is actually a slight tendency for the sign-score to work "backwards" at the lower end. We would not anticipate this result if all eight signs were used, but at the moment there is no evidence for this disclaimer.

Judging from the cross-validational evidence the method of profile interpretation advocated here does possess validity for predicting success in practice teaching. Its practical efficiency may not be high, but the empirical confirmation of the methodology indicates that a systematic and diligent search for signs and patterns might well yield predictors of practical utility.[8] Certainly the results highlight the error involved in concluding that the MMPI does not "work" because single scale analysis fails.

The predictive efficiency observed in this study is also, of course, attenuated by the unreliability of the criterion. It has not been possible to evaluate the ratings systematically, but the impression is gained that ratings vary considerably in accuracy and value. A further study such as the present one would want to insure the obtaining of a valid and reliable set of criterion measures if a goal of practical forecasting efficiency is to be met. There is also the possibility that better results could be achieved by differentiating the criterion—performance in practice teaching—into its constituent phases, instead of treating it in an over-all way.

Table 5

Predictive Efficiency of Eight MMPI "Signs" in the Original Sample

		Practice Teaching Ratings			
		Higher Ratings	Average Ratings	Lower Ratings	
Number of Signs	6–8	12	16	1	29
	3–5	19	46	22	87
	0–2	5	13	13	31
		36	75	36	147

$\chi^2 = 14.29$; d.f. = 4; and $P < .01$.

Table 6

Cross-validational Predictive Efficiency of Five MMPI "Signs"

		Practice Teaching Ratings		
		Higher Ratings	Lower Ratings	
Number of Signs	4–5	27	15	42
	2–3	39	54	93
	0–1	14	11	25
		80	80	160

$\chi^2 = 6.21$; d.f. = 2; and $P > .02$, $< .05$.

Summary

An attempt was made to predict success in practice teaching from personality test measures. Single scales on the MMPI showed little validity but various patterns and indices revealed considerable promise. Certain methodological implications of this study for projects devoted to goals of practical assessment and evaluation were discussed.

Received September 28, 1951.

References

1. Gough, H. G. A new dimension of status: I. Development of a personality scale. *Amer. sociol. Rev.*, 1948, **13**, 401–409.
2. Gough, H. G., McClosky, H., and Meehl, P. E. A personality scale for dominance. *J. abnorm. soc. Psychol.*, 1951, **46**, 360–366.
3. Hathaway, S. R., and McKinley, J. C. *The Minnesota multiphasic personality inventory.* Minneapolis: University of Minnesota Press, 1943.
4. Pemberton, W. H. *Test characteristics of student teachers rated at the extremes of teaching ability.* Doctoral dissertation, Univ. of Calif., Berkeley, 1950.
5. Welsh, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16**, 65–72.

[8] One technique for searching for signs would be to apply Meehl's configural method for *items* (see Meehl, P. E. On configural scoring. *J. consult. Psychol.*, 1950, **14**, 165–171) to pairs of *scales*. A current investigation seeking to differentiate patients with ulcers from those with other diagnoses has achieved encouraging results utilizing this modification. (See Sullivan, P. L., and Welsh, G. S. A technique for objective configural analysis of MMPI profiles. *J. consult. Psychol.*, 1952, **16**, 383–388.)

If a study similar to the present one were to be carried out on a large scale ($N \geq 500$ in both original and cross-validating samples) it would undoubtedly be advantageous to combine the straightforward empirical technique advanced by Meehl with the somewhat more intuitive mode of discovery employed here. In either case, however, the *validation* of the sign should rest squarely upon demonstrable relationships and should not be left in the realm of speculation and conjecture.

Predictive Value of The Empathy Test in Automobile Salesmanship

Francis P. Tobolski and Willard A. Kerr

Illinois Institute of Technology

Examination of the literature on job success in various fields of salesmanship reveals that empathic ability measurement has not been investigated as a possible useful predictor (**1, 3, 6**). In fact, a standardized and validated objective test of empathic ability was not available until 1951 (**4, 5**). The absence of empathic ability measurements in salesmanship prediction batteries seems a remarkable historical oversight by psychologists. On the surface at least, it appears fairly certain that the success of the salesman usually depends significantly on his ability to assimilate and anticipate the feelings of others. This study tests that hypothesis for the job of selling automobiles.

Subjects and Procedure

Subjects for this study were sales personnel of two of Chicago's largest automobile agencies. Twenty-two (seven new car and fifteen used car salesmen) of the Company A sales personnel of twenty-five and all ten (four new and six used car salesmen) of the Company B salesmen participated in the study.

The instrument used for measuring empathic ability was an approximately fifteen-minute, completely objective test, *The Empathy Test* (**4**). Its rationale, explained elsewhere (**4, 5**), hypothesizes that the presence of empathic ability may be demonstrated by the ability to predict representative behaviors of normative individuals. *The Empathy Test* applies this rationale by requesting the respondent to predict the responses of typical individuals in key behavior areas—aesthetics (music types), general human interests (what people read), and interpersonal relations (annoying experiences).

Criteria for job success were sales ratio (cars sold divided by attempts) over a three-month to twelve-month period and sales managers' rank order ratings. Four validity coefficients were computed on each of the two criteria since each company included salesmen specializing separately in used car or new car selling.

Results

Inspection of Table 1 shows a mean (Fisher z-transformed) rank difference coefficient of validity of .44 against the sales record criterion. The mean validity coefficient against the sales manager rankings is .71. Both of these coefficients are significant at the 98 probability level, or better, according to the Thornton (**7**) tables. Approximately one-fifth of the variance in objective sales records is accounted for by variance in scores on *The Empathy Test*.

An interesting result is the consistent tendency for validation coefficients on the salesmen of new cars to exceed those of the salesmen of used cars. It is difficult to account for this difference except, perhaps, as an artifact of the differential effects of television advertising of new and used cars on the actual final selling situation. Both companies employ television as their principal medium of advertising. The television viewer who is interested in the new

Table 1

A Summary of Validity Coefficients from *The Empathy Test* on Groups of Salesmen Employed by Two Leading Chicago Auto Dealers

Salesmen Group	N	Rank Difference Validity Coefficients*	
		Sales Record	Ranking by Sales Manager
Company A			
New car	7	.42	.50
Used car	15	.05	.13
Total	22	.25	.33
Company B			
New car	4	.75	1.00
Used car	6	.18	.20
Total	10	.58	**.89**
New car total	11	.61	**.92**
Used car total	21	.12	.17
Total both companies	32	**.44**	**.71**

* Coefficients which are reliable at the 95 probability level are indicated in boldface.

cars finds them as typically represented upon visiting salesrooms, but the television viewer interested in used cars finds that the typical vehicles in the used car lots may differ considerably from their glossy appearance on television. Theoretically, the potential used car customer would therefore experience greater frustration and disillusionment than does the potential new car customer when he examines the actual vehicles. Aggression from this greater frustration (2) of the potential used car customer seems most likely to be channeled against the nearest plausible object, which is, of course, the used car saleman. Thus the used car salesman may find that this customer aggression factor prevents as consistent a sales performance as is obtained by a new car salesman of equal *empathic* ability. It is possible that this frustration factor results in the potential customer largely discounting sales contentions and buying largely on the basis of other evidence and observations.

Summary

Thirty-two salesmen of two of Chicago's largest new-and-used automobile agencies were examined on *The Empathy Test* and their scores were correlated with two criteria.

1. *Empathy Test* scores were significant predictors of sales records ($r = .44$).

2. *Empathy Test* scores were significant predictors of the job success of sales crew members as ranked by their sales managers ($r = .71$).

3. *Empathy Test* score predictions of sales and success ranking of used car salesmen were very inefficient (r's = .12 and .17).

4. The superior predictive value of the test on new car salesmen is tentatively attributed to the differential effects of television advertising on the selling operation for new and used cars. A frustration-aggression situation is hypothesized to arise more frequently in the latter.

Received October 24, 1951.

References

1. Buros, O. K. *Third mental measurements yearbook.* New Brunswick, New Jersey: Rutgers University Press, 1949.
2. Dollard, J. L., Doob, N. J., Miller, N. E., Mowrer, O. H., and Sears, R. R. *Frustration and aggression.* New Haven: Yale University Press, 1939.
3. Dorcus, R. M., and Jones, M. H. *Handbook of employee selection.* New York: McGraw-Hill Book Co., Inc., 1950.
4. Kerr, W. A., and Speroff, B. J. *The empathy test.* Chicago 90: Psychometric Affiliates, 1951.
5. Speroff, B. J., and Kerr, W. A. *Validation and evaluation of the empathy test.* Presented at annual meeting, Midwestern Psychological Association, Drake Hotel, Chicago, April 27, 1951.
6. Super, D. E. *Appraising vocational fitness.* New York: Harper and Bros., 1949.
7. Thornton, G. R. The significance of rank difference coefficients of correlation. *Psychometrika,* 1943, **8**, 211–222.

Predicting Success of Students in Veterinary Medicine*

Wilbur L. Layton

Student Counseling Bureau, University of Minnesota

The enrollment in schools of veterinary medicine has increased by approximately 500 per cent in the last twenty-five years (**2**). The competition for admission to these schools has been almost as great as that for admission to medical schools. Counselors, teachers and administrators therefore must have available indices predictive of the success or failure of students in veterinary medicine. These indices are potentially useful for the selection of students and for the counseling of students, before and after their admission to schools.

Recently, a Veterinarian Aptitude Test was devised by Owens and Payne of the Iowa State College (**4**, **5**) and Hannum developed a veterinary medicine scale for the Strong Vocational Interest Blank (**1**). The present study was designed to evaluate these and other tests and predictive indices forecasting the academic achievement of students enrolled in the School of Veterinary Medicine at the University of Minnesota.

Method

In the fall quarters of 1948 and 1949, freshmen enrolled in the School of Veterinary Medicine at the University of Minnesota were tested on the Strong Vocational Interest Blank, the Iowa State College Veterinary Aptitude Test, and two parts of the Professional Aptitude Test of the Educational Testing Service. The Iowa Test consists of three parts: reading comprehension, verbal memory, and preveterinary achievement in chemistry and zoology. The two parts of the professional aptitude test used were the following: the verbal ability section, yielding verbal ability, scientific; verbal ability, special; verbal ability, humanistic and composite verbal ability scores, and the pre-medical science achievement section.

Grades in chemistry and other physical science courses taken as a part of the preveterinary work were available and combined as one predictive variable. Pre-veterinary grades in botany and zoology were combined in a similar manner. Grades in all preveterinary medicine courses were combined as a third measure of pre-veterinary achievement. The ages of the students were also available.

A total of 100 students were admitted as freshmen in the two classes studied. Complete data were available for 87 males. When grades in the School of Veterinary Medicine became available, data for the two classes were combined and the relationships determined between the predictive measures and cumulative honor-point-ratios at the end of the first year in the School of Veterinary Medicine. The Strong Vocational Interest data were analyzed first, then selected scales on the Strong Blank and the other predictive indices were correlated with the criterion, first year grades, and with each other. Several combinations of these variables were tried out to determine their predictive efficiency. One combination was selected as giving optimal prediction.

Analysis of the Strong Vocational Interest Blank

Distributions of the letter ratings on 39 occupational scales and three non-occupational scales for the Strong Vocational Interest Blank were dichotomized and related to the criterion, first year grades. Chi-square tests were run to determine statistically significant relationships. The following six scales were found to be related to the criterion with Chi-squares significant at the one per cent level: osteopath, chemist, farmer, aviator, Y.M.C.A. secretary, and sales manager.

The veterinarian scale (**1**) became available when this study was almost completed. The

* The statistical analysis for this study was performed by the Bureau of Institutional Research, University of Minnesota. Appreciation is expressed to Dr. Robert Keller, Director, and Margaret Abernathy, Research Fellow, Bureau of Institutional Research, and to Professor Martin H. Roepke of the University of Minnesota School of Veterinary Medicine for their assistance and cooperation in this study.

Table 1

Coefficients of Correlation Between Predictive Indices and Freshman HPR for Classes (N = 87) Entering the School of Veterinary Medicine in 1948 and 1949

Predictor	Coefficient of Correlation	Mean	Standard Deviation
Iowa State Veterinary Aptitude Test			
1. Paragraph Comprehension	.21	49.28	5.57
2. Verbal Memory	.03	23.40	4.01
3. Pre-veterinary Achievement	.27	57.53	6.99
4. Total (1+2+3, unweighted)	.25	130.21	12.16
Professional Aptitude Test			
5. Verbal Ability, Scientific	.22	29.95	9.83
6. Verbal Ability, Special	.00	25.61	9.33
7. Verbal Ability, Humanistic	.03	16.63	8.17
8. Verbal Ability, Total	.10	72.20	23.52
9. Pre-medical Science	.30	46.28	17.19
Strong Vocational Interest Blank			
10. Osteopath	.28	43.52	8.95
11. Veterinarian	.30	47.22	8.17
12. Chemist	.22	34.33	8.73
13. Farmer	.12	50.38	6.84
14. Aviator	.28	40.80	8.04
15. Y.M.C.A. Secretary	.12	23.79	10.29
16. Sales Manager	.15	23.94	6.74
Pre-Veterinary Grades*			
17. Chemistry HPR	.42	1.87	.54
18. Zoology and Botany HPR	.43	2.05	.51
19. Total Pre-Vet. HPR	.53	1.94	.31
20. Age at Entrance to Vet. Med.	−.33	23.89	3.15
21. Freshman HPR	—	1.73	.56

* Expressed as an honor-point-ratio (honor-points divided by number of credits). Honor points were assigned as follows: A = 3, B = 2, C = 1, D, F = 0.

Strong answer sheets were rescored for it and it was analyzed separately by the above method. A significant relationship (one per cent level) was found between it and the criterion.

The first six scales listed above were included in the entire correlation study. The veterinarian scale was included in the final stages of the correlation study.

Table 2

Intercorrelations of the Variables Used in Trial Multiple Correlations

Variable	Strong Osteo. 10	Strong Aviator 14	Age 20	P.A.T. Pre-Med. Sci. 9	Ia. Pre-Vet. Ach. 3	Pre-Vet. HPR 19	Strong Vet. 11
21	.28	.28	−.33	.30	.27	.53	.30
10		.31	−.20	.15	.17	.28	.27
14			.22	−.03	.18	.06	.08
20				−.05	−.19	−.27	−.12
9					.65	.23	.14
3						.21	−.01
19							.30

The Correlation Study

The raw scores or standard scores for each of the variables were converted into stanine scores. These stanines were punched into IBM cards and a table of intercorrelations (Pearsonian) was computed. Table 1 presents the validity coefficients, mean raw or standard scores and standard deviations for all variables.

Selected variables were tried in combination to determine their predictive efficiency. Table 2 indicates the intercorrelations of the variables used in the various combinations. Multiple correlation coefficients were computed for these several combinations. These correlations, taken in pairs, were compared using the F test. Table 3 presents the various combinations tried, the resulting multiple correlation coefficients, and results of the F tests. The last combination of variables listed, with the fewest number of variables, yielded a coefficient of multiple correlation which compared well with the larger ones and hence appeared best for practical use.

The three variables combined in the final regression equation to predict first year grades were: the veterinarian scale on the Strong, pre-veterinary total honor-point-ratio, and pre-veterinary achievement score on the Iowa Veterinary Aptitude test.

The regression equation is:

$$y = .86 + .43\,X_1 + .23\,X_2 + .17\,X_3$$

Table 3

Combinations of Variables with Resulting Coefficients of Multiple Correlation and Levels of Significance of the F Tests of the Difference Between Pairs of Multiple Correlations

Variables in Combination	Coefficient of Multiple Correlation	Level of Significance of the F Test
10, 14, 20, 9, 19, 11	.6537	<.05
10, 14, 20, 9, 19, 3, 11	.6678	
10, 14, 20, 3, 19, 11	.6506	<.05
10, 14, 20, 9, 3, 19, 11	.6678	
10, 14, 20, 9, 3, 19, 11	.6506	.05
10, 3, 19, 11	.6134	
10, 14, 20, 3, 19, 11	.6506	.05
3, 19, 11	.6017	

Table 4

Conversion of Scores to Stanines

Freshman Year H.P.R.*	Total Pre-Vet. H.P.R.*	Strong Vet. Scale Standard Scores	Iowa Pre-Vet. Achievement R.S.	Stanine
3.0	2.5–2.8	61–72	68–70	9
2.5–2.9	2.3–2.4	57–59	66–67	8
2.1–2.4	2.1–2.2	52–56	63–65	7
1.8–2.0	1.9–2.0	50–51	60–62	6
1.5–1.7	1.7–1.8	44–49	57–59	5
1.3–1.4	—	42–43	52–56	4
1.1–1.2	1.6	36–41	47–51	3
1.0	1.5	32–35	44–46	2
.4–.9	1.1–1.3	28–31	40–42	1

* Expressed as an honor-point-ratio (honor-points divided by number of credits). Honor points were assigned as follows: A = 3, B = 2, C = 1, D, F = 0.

where

y = Predicted honor-point-ratio, first year in Vet. Med.
X_1 = Total honor-point-ratio in pre-veterinary courses.
X_2 = Veterinarian scale on the Strong.
X_3 = Pre-veterinary achievement test score of the Iowa Test.
$R = .60$.

y, X_1, X_2, and X_3 are expressed in stanines. Table 4 presents the conversions of raw or standard scores into stanines for these four variables.

Discussion

The correlation of .53 between total pre-veterinary H.P.R. and freshman grades is somewhat higher than that reported by Owens (3). However, the correlations found for chemistry H.P.R. and zoology and botany H.P.R. are almost identical with those reported by him. Owens reported relatively high validity coefficients (r's ranging from .48 to .72) for the Veterinary Aptitude Test. In the present study neither aptitude test correlated highly with the criterion. In fact the selected interest scales correlated about as highly.

This is one of the few instances known to the author in which a scale on the Strong has made a significant contribution to a regression equation predicting students' grades. The School

of Veterinary Medicine at the University of Minnesota is a newly established one. The two classes studied were the second and third classes admitted to the school. The faculty has given a great deal of personal attention to these students. One might hypothesize that this situation is responsible for the relationship between interests and grades.

Summary

Twenty variables were studied to determine their effectiveness in predicting grades earned by first year students in veterinary medicine. Total honor-point-ratio in pre-veterinary course work, the veterinarian scale on the Strong Vocational Interest Blank and the pre-veterinary achievement test of the Iowa State College Veterinary Aptitude test were combined in the final regression equation and yielded a coefficient of multiple correlation of .60.

Received July 7, 1952.
Early publication.

References

1. Hannum, T. E. Response of veterinarians to the Strong Vocational Interest Blank for Men. *Proc. Ia. Acad. Sci.*, 1950, **57**, 381–384.
2. Klussendorf, R. C. Education in veterinary medicine. *Higher Educ.*, 1949, **5**, 181–185.
3. Owens, W. A. Development of a test of aptitude for veterinary medicine. *Proc. Ia. Acad. Sci.*, 1950, **57**, 417–423.
4. Owens, W. A., and Payne, L. C. *The Iowa State College Veterinary Aptitude Test.* Ames, Iowa, 1950.
5. Owens, W. A. An aptitude test for veterinary medicine. *J. appl. Psychol.*, 1950, **34**, 295–299.

The Criterion Problem in the Prediction of Medical School Success*

Eugene L. Gaier

University of Illinois

The current study was initiated for the purpose of investigating the criterion used in the professional medical training program. The specific purpose of the present paper is to review the procedures used to evaluate performance in one medical school and to determine the adequacy of these procedures in terms of the objectives of medical education both from a logical and a statistical point of view.

A basic definition here is that of a criterion as a set of symbols used to describe the performance of individuals on a success continuum (4). The measure of success emphasized in medical training has been the average of four years' marks on the assumption that the marks or grades assigned represent the extent to which the objectives of the medical school have been achieved. The characteristics of such a criterion might be grouped for purposes of discussion as: (1) relevance or pertinence; (2) comprehensiveness; (3) reliability and discrimination; and (4) freedom from bias. The first two of these characteristics involve judgment as to the adequacy with which the proposed *success measures* represent performances on the defined dimension of success. The third and fourth categories deal with the more formal aspects common to any set of psychological measuring techniques. These aspects are evaluated in terms of the assigned marks as indicators of the level of performance.

Procedure

The records of the students of two classes who entered the Medical College of the State University of Iowa were used as the data for this investigation. They were selected as being two representative classes of this school, and will henceforth be referred to as Class I and Class II.

Class I. School records for ninety individuals, 5 women and 85 men, were available. The ages ranged from 20 to 34. Thirty-five members (38.8 per cent) of this class completed all of their premedical training at the State University of Iowa, while 26 (28.8 per cent) of the students had done part of their premedical training at this institution. Sixty-six cases (73.3 per cent) completed the four year medical program and received the M.D. degree.

Class II. Records for this class were available on 4 women and 83 men, for a total of 87 cases. Nineteen members (21.9 per cent) of this group completed some of their premedical education at the State University of Iowa, while 38 (43.6 per cent) of the members completed all of their premedical training at this institution. Seventy-four (84 per cent) of the students completed the four-year curriculum and received the M.D. degree.

For each student, data on the following items were punched on IBM cards: age, sex, premedical school, premedical grade point average,[1] national percentile rank on the Moss Aptitude Test,[2] year and reason for leaving medical school, the letter grade received for each of the 42 courses in the four-year medical curriculum, and the grade point average for each year of medical work.

The undergraduate course record as well as the four yearly weighted grade point averages were available for each student. Credits of four, three, two, one, and zero per unit hour were assigned to grades A, B, C, D, and E, respectively, in determining a total grade point for each student. Each grade was assigned an additional weight in terms of the number of clock hours in the course.

Results

The analysis of the data on the criterion measures will be presented under the following

* This study is a condensation of a Master's thesis submitted at the State University of Iowa. The author wishes to express appreciation to Dr. Harold P. Bechtoldt under whose direction this study evolved.

[1] This coded premedical grade point average will henceforth be called the premedical average grade.

[2] The coded national percentile on the Moss Aptitude Test will henceforth be called the Moss Aptitude Scores.

four headings: (1) frequency distributions of grades; (2) intercorrelations of eleven selected variables; (3) reliability of the assigned grades; and (4) prediction of medical school success.

1. Frequency Distribution of Grades. Frequency and percentage distributions of course grades for the two classes are presented in Table 1. For both classes, the distributions of grades given in the last three years of work are very similar ($\chi^2 = 6.61$, $P > .10$ for the two combined distributions, with four degrees of freedom). The two distributions of grades given during the freshman year, however, show differences significant at the one per cent level ($\chi^2 = 49.14$, $P < .01$ with four degrees of freedom).

While Class II had fewer students eliminated, the difference in the proportions securing degrees in the two classes (73.8 per cent and 85 per cent, respectively) is not significant at the five per cent level of confidence. The similarity of the grade distributions for the last three years together with the insignificant difference in the number of failures in the two classes during those years suggest either that the performances of the two classes over the four years are comparable or that the standards of grading may shift from year to year to accommodate differences in general performance level. Although these alternative explanations cannot be directly evaluated without the use of comparable, objective achievement tests, some information can be secured from the differences in the mean scores on the predictor variables and the average grades assigned during the first year.

The differences between the means of the two samples were significant at the one per cent level ($t \geqq 2.58$, degrees of freedom > 100) for the following four variables: anatomy, histology, the premedical average grade, and the Moss test scores. The differences in the means[3] of the Moss test and the premedical grade point average can be considered as representing a difference in initial performance in the factors represented by these variables. In so far as the scores on the Moss and the premedical grade averages represent compar-

[3] The assumption of homogeneity of variances required by the tests of significance of differences in means was established for all variables over the two samples at better than the five per cent level of confidence.

Table 1

Frequency and Percentage Distributions* of Letter Grades Assigned in All Courses for the Four Year Course in the Medical College of the State University of Iowa (Classes I and II)

Letter Grade	First Year				Second Year				Third Year				Fourth Year				Grand Total			
Class:	I		II		I		II		I		II		I		II		I		II	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
A	16	3.3	25	4.9	66	8.1	78	8.2	86	10.0	74	7.4	100	15.1	91	11.2	268	9.5	268	8.2
B	101	21.2	192	37.8	213	26.1	275	29.0	304	35.4	309	30.9	24	37.6	341	41.9	867	30.8	1,117	34.2
C	258	54.2	237	46.7	434	53.2	507	53.5	390	45.5	497	49.7	282	42.6	314	38.6	1,364	48.5	1,555	47.6
D	72	15.2	43	8.5	98	12.0	88	9.2	78	9.1	111	11.1	31	4.7	66	8.1	279	9.9	308	9.4
E†	29	6.0	10	2.0	5	0.6	0	0.0	0	0.0	9	0.9	0	0.0	2	0.2	34	1.2	21	0.6
N =	476		507		816		948		858		1,000		662		814		2,813		3,269	

* The number of letter grades assigned will vary from year to year dependent on the number of courses included in the schedule for the year, and the number of students enrolled.

† A letter grade of "E" represents a failing mark.

able standards of evaluation for the two classes, Class II can be considered the more promising of the two groups.

If only the less capable students have been rejected, then the premedical averages and the Moss scores (see Tables 2 and 3) of the students receiving the M.D. degree should not differ significantly. The differences between the mean values of these two variables are significant at better than the one per cent level of confidence. The groups receiving the M.D. degree, therefore, cannot be considered homogeneous with respect to mean scores on the Moss and the premedical grade point averages. Furthermore, for these selected students, the correlation coefficients between the two selection variables and grades received in the five courses of the freshman year are not significantly different for the two samples, 16 of the 20 coefficients being significantly different from zero at the five per cent level ($N = 66$, $r = .24$). One of the four non-significant correlations ($r = .19$) was between the Moss and the histology grades for Class II. It would seem reasonable to conclude here that the factors represented by the two selection variables were about equally important for success in the five freshman courses. Since the significant differences in course grades were confined to two courses and since the significant differences in the selection variables were not eliminated by rejecting failing students, the grading procedures apparently were not operating to eliminate the students with the lowest standings on these variables.

Inasmuch as the only significant difference in the average grades assigned the two groups for each of the four years was for the first year average, the data are considered as supporting the hypothesis that these differences in initial ability were not responsible for the differences in the mean scores of the five first year courses. In other words, the results are interpreted as indicating that the standards of evaluation used for the two groups probably were not strictly comparable.

2. Intercorrelation of the Variables. The complete correlational analysis was restricted to eleven variables: five freshman courses,[4] the average grades for each of the four years of medical school, and the two predictor variables. The data on five of the six freshman courses were included to provide an estimate of the reliability of the first year grades in predicting relative success in the last three years of medical school.

Thirty-seven of the correlations for Class I and 42 correlations for Class II (50 correlations were computed for each class) were significantly different from zero at the one per cent level.[5] Since the correlations were all zero or positive, the hypothesis that a single, common factor would account for the intercorrelations[6] of ten variables, excluding first year average grades, was investigated. The hypothesis was tested in terms of the magnitude of the difference between the Z's corresponding to the observed and predicted correlations. The predicted correlations were computed from the cross product loadings on the single common factor by the formula $r_{jk} = a_j a_k$. Since this procedure dealt with the discrepancy between a theoretical and observed value of a Z, the standard error of the theoretical Z was defined as $\sigma_Z = \frac{1}{\sqrt{N-3}}$.

Significant residuals between the observed and theoretical Z's (five per cent level) were found for three pairs of variables: (1) second vs. third year average grades; (2) anatomy vs. physiology grades; and (3) biochemistry vs. physiology. Only one residual corresponding

[4] Detailed analysis of separate course grades could be justified only for five of the six freshman courses; these courses accounted for eighty-five and fifty-eight per cent of all the failing marks assigned to Class I and II, respectively.

[5] The correlations which were not significant from zero although at the five per cent level for Class I were: physiology with anatomy, fourth year average with anatomy, premedical grade point with anatomy, fourth year average with histology, premedical grade point with histology, neuroanatomy with physiology, neuroanatomy with second year average, neuroanatomy with third year average, neuroanatomy with premedical grade point, neuroanatomy with the Moss test, second year average with the Moss test, third year average with the Moss, and the fourth year average with the Moss. Non-significant correlations for Class II were: anatomy with fourth year average, anatomy with the Moss, histology with premedical grade point, histology with the Moss, biochemistry with the Moss, physiology with the Moss, neuroanatomy with the Moss, the fourth year average with premedical grade point average, and fourth year average with the Moss test.

[6] The factor loadings were computed by Thurstone's direct summation method for the special case of unit rank (5, p. 227).

Table 2

Means, Standard Deviations, and Intercorrelations for Eleven Variables for the Iowa Medical College Class I (N = 66)

Code	Variable	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11
1	Anatomy—first year	2.2	.51	—	.57	.42	.16	.46	(.76)	.49	.34	.25	.23	.32
2	Histology—first year	2.1	.57	.57	—	.48	.35	.36	(.72)	.50	.34	.30	.27	.44
3	Biochemistry—first year	2.1	.83	.42	.48	—	.58	.50	(.82)	.50	.48	.40	.52	.32
4	Physiology—first year	2.0	.65	.16	.35	.58	—	.21	(.67)	.28	.24	.32	.24	.28
5	Neuroanatomy—first year	2.4	.67	.46	.36	.50	.21	—	(.57)	.54	.50	.43	.38	.22
6*	First year average grade (six courses)	2.1	.46	(.76)§	(.72)	(.82)	(.67)	(.57)	—	.61	.49	.41	.45	.39
7*	Second year average grade	2.2	.42	.50	.50	.50	.28	.54	.61	—	.71	.34	.32	.29
8*	Third year average grade	2.4	.45	.34	.34	.48	.24	.50	.69	.71	—	.47	.35	.28
9*	Fourth year average grade	2.6	.32	.25	.30	.40	.32	.43	.41	.34	.47	—	.43	.21
10†	Premedical g.p. average	3.5	1.78	.23	.27	.52	.24	.38	.45	.32	.35	.43	—	.38
11‡	Percentile rank on Moss Aptitude Test	5.2	2.51	.32	.44	.32	.28	.33	.39	.29	.28	.21	.38	—

* The average grade represents a weighted average based on the number of clock hours per course and an assigned value for each letter grade.
† The premedical grade point averages coded as single digit scores by a linear transformation.
‡ The national percentile ranks on the Moss Aptitude Test coded as single digit scores by a linear transformation.
§ The correlations in parentheses are not experimentally independent and should be disregarded.

to the correlation between second and third year average grades was significant at this level of confidence for Class II. These results neither clearly sustain nor refute the hypothesis for a single factor; the results can be considered equivocal.

Since the hypothesis of a single factor is not clearly sustained, the possible influence of changes in the course content on the pattern of intercorrelations (see Tables 2 and 3) of the average grades might account for other factors. The low correlations between the grades in the last two years may reflect the unreliability of judgments of performance in clinical work. However, the correlations of .71 and .75 between the second and third year grades would indicate that the reliability of the third year clinical ratings was perhaps comparable to that of the second year marks. Therefore, if low reliability is the major factor in the low correlations of the last two years, the unreliability would appear to be concentrated in the grades of the senior year.

Although the pattern of intercorrelations among the average grades for the four years is not consistent with the predicted values, the fact that the correlations are positive and significant does indicate a degree of consistency necessary for an acceptable criterion. The series of grades included here, in fact, might be considered as arising from the presence of a single, common factor. Further support for the hypothesis is provided by the fact that the relations are consistent for both samples studied.

3. Reliability of the Assigned Grades. The reliability of a composite first year grade, defined as the simple sum of grades in five courses was estimated on the assumption that a single, general factor was sufficient to account for the intercorrelations of the grades. This assumption was tested: (1) by determining the factor loadings for the special case of unit rank[7] from the intercorrelation of the five

[7] Thurstone's (5) direct summation method of evaluation was used.

Table 3

Means, Standard Deviations, and Intercorrelations for Eleven Variables for the Iowa Medical College Class II (N = 74)

Code	Variable	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11
1	Anatomy—first year	2.7	.62	—	.32	.51	.44	.45	(.82)	.47	.42	.28	.45	.27
2	Histology—first year	2.8	.49	.32	—	.33	.33	.32	(.50)	.30	.37	.35	.25	.19
3	Biochemistry—first year	2.3	.78	.51	.33	—	.64	.53	(.81)	.53	.46	.48	.44	.21
4	Physiology—first year	2.3	.78	.44	.33	.64	—	.49	(.80)	.58	.43	.44	.31	.24
5	Neuroanatomy—first year	2.4	.64	.45	.32	.53	.49	—	(.64)	.55	.50	.43	.33	.11
6*	First year average grade (six courses)	2.5	.51	(.82)§	(.50)	(.81)	(.80)	(.64)	—	.64	.55	.49	.37	.37
7*	Second year average grade	2.3	.40	.47	.30	.53	.58	.55	.64	—	.75	.33	.44	.43
8*	Third year average grade	2.4	.42	.42	.37	.46	.43	.50	.55	.75	—	.47	.43	.42
9*	Fourth year average grade	2.4	.32	.28	.35	.48	.44	.43	.49	.33	.47	—	.20	.19
10†	Premedical g.p. average	4.3	1.59	.45	.25	.44	.31	.33	.37	.44	.43	.20	—	.35
11‡	Percentile rank on Moss Aptitude Test	5.9	2.23	.27	.19	.21	.24	.11	.37	.43	.42	.19	.35	—

* The average grade represents a weighted average based on the number of clock hours per course and an assigned value for each letter grade.
† The premedical grade point averages coded as single digit scores by a linear transformation.
‡ The national percentile ranks on the Moss Aptitude Test coded as single digit scores by a linear transformation.
§ The correlations in parentheses are not experimentally independent and should be disregarded.

freshman courses; and (2) by determining the significance of the difference between the observed and predicted Z values. None of the differences was significant at the five per cent level; the assumption, therefore, could be accepted.

The reliability coefficient of the grades assigned was estimated by two procedures. In the first estimate, the analysis of variance method formulated by Burt (**1**) and Hoyt (**2**) was employed. The error variance is defined as the interaction of residual variance. Since the variances between individuals were significant for both sets of data beyond the one per cent level ($F_1 = 18.48$ and 23.38 for Classes I and II, respectively),[8] the reliability coefficient can be considered significant. Both of these reliability coefficients were estimated[9] as .95. This procedure also provided a test of the significance between the means of courses for each group ($F_2 = 4.94$ and 12.88 for Classes I and II, respectively).[10] Both F ratios were significant at beyond the one per cent level, and the variations between mean grades of the courses for each group are greater than those expected by chance.

The second estimate of reliability was secured by the Kuder-Richardson formulation (**3**). The highest of the four intercorrelations of the freshman course grades should provide a conservative estimate of the reliability of the separate grades. By this method, the reliability coefficients for the averages were .89 and .90 for Class I and Class II, respectively. It can be concluded, therefore, that the reliability coefficient probably lies between .89 and .95 for the average of the first year grades, a value that indicates a satsifactory degree of stability.

[8] F_1 is defined as the variance between individuals divided by the variance for interaction.

[9] r_{ii} is defined as the ratio of the difference between the variance between individuals and the error variance to the error variance.

[10] F_2 is defined as the variance between courses divided by the error variance.

The resulting lower intercorrelations between the first year average grades and the other variables could not be attributed to the unreliability of these marks. It may be concluded, therefore, that the first year marks do exhibit a desirable degree of discrimination and reliability. The obtained correlations (.61 and .64) of first and second year grades are significantly lower than those reliability coefficients. The attenuation of these correlations is probably not due to the unreliability of the marks, but rather to factors such as the shift of the objectives in the course instruction or in the measurement of these objectives from instructor to instructor.

4. Prediction of Medical School Success. The preceding analysis has indicated that the criterion measures may be considered as having some definite factor in common and that the reliability and the discrimination capacity of the first year grades were adequate. The relationship of two possible predictor variables to these measures of medical school success was investigated. The selection variables used were the premedical grade point average, and the percentile rank on the Moss Aptitude Test. The analysis included the computation of the zero-order correlations and the multiple correlations for all cases in each class receiving the M.D. degree. Point biserial correlations were determined for all members of the freshman class; the criterion was "pass-fail" in medical school.

The zero-order correlations between the premedical grades and the four criterion measures (i.e., first, second, third, and fourth years of medical averages for the selected cases) are contained in Tables 2 and 3. For Class I, the highest correlation of .45 with the above criteria was obtained between the premedical grade point average and the first year average grade in medicine. The correlations of premedical grades with the second and third year average were higher for Class II than for Class I (.44 and .43 for Class II second and third year grades). The differences in the correlations for the two classes, however, are not statistically significant as noted above.

The estimated multiple correlation coefficient between the criterion measures and combined scores on the Moss and the premedical average exceed the larger zero-order correlations by at least .05 for five of the eight coefficients. For Class I, the multiple correlations for the first three years are significant, at the one per cent level. Prediction of fourth year achievement by the multiple correlation technique was not improved over that from the zero-order correlation value of .43. The combined use of the Moss and premedical grade point may be recommended when the selection ratio employed is low, but the effectiveness of the prediction of rank in class is low. Average grades for Class II can be predicted more closely by a combination of the two predictor variables during the second and third years; and on every level, a combination of the two variables results in a greater value than the zero-order correlation by itself

Summary

This study has been concerned with the characteristics of the measures employed as the criterion of success in one medical school. With respect to the formal aspects of the criterion, the findings may be summarized as follows:

1. The consistency of the distributions of grades for the two classes were not of comparable ability and achievement. The conclusions reached were that either (a) the differences in the measure of ability were reflected in, or pertinent to, only two of the freshman courses; or (b) the standards of evaluations used varied from class to class. The second of these alternatives was considered more tenable.

2. That the intercorrelations found between eight measures of medical performance (grades) and two measures of premedical achievement (premedical average grades and the Moss Aptitude Test) could be accounted for by a single variable or common factor was equivocal, since these differences between the observed and predicted Z values were significant at the five per cent level.

3. The hypothesis of a single common factor for the five freshman courses was sustained. These five courses can then be considered as measuring some single hypothetical variable, and would be interpreted as a complex function of ability, motivation, and work habits.

4. On the assumption of a single, common factor, two different estimates of the reliability coefficient of the average course grades were obtained: .89 and .95 for Classes I and II, respectively. The more conservative estimates were based on the Kuder-Richardson formulation with the separate course reliabilities estimated as the highest intercorrelations. The higher estimate was secured from an analysis of variance procedure. These estimates indicated that the freshman grades possess a desirable degree of internal consistency.

5. The relations between relative standings on two measures of ability (premedical grade point average and the Moss Aptitude Test) and success in the freshman courses were generally low but of comparable magnitude.

Received October 17, 1951.

References

1. Burt, C. The reliability of teachers' assessments of their pupils. *Brit. J. educ. Psychol.*, 1945, **15**, 80–92.
2. Hoyt, C. Test reliability attained by analysis of variance. *Psychometrika*, 1941, **6**, 153–160.
3. Kuder, G. F., and Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, **2**, 151–160.
4. Stuit, D. B. *Personnel research and test development.* Princeton: Princeton Publishing Co., 1947.
5. Thurstone, L. L. *Multiple factor analysis.* Chicago: Univ. of Chicago Press, 1947.

A Comparison of Three Criteria of Aircrew Effectiveness in Combat Over Korea*

John K. Hemphill and Lee B. Sechrest

Personnel Research Board, The Ohio State University

This paper reports a study of three criteria of the performance of 94 B-29 aircrews which flew combat missions over Korea during the period extending from March to September 1951. The three criteria to be considered are: (1) ratings by superiors of the performance of crews as units; (2) sociometric nominations from crew members; and (3) objective records of combat bombing accuracy. Following a brief description of each of these criteria, they will be compared in terms of their reliability and interrelationship. A general problem in the use of superiors' ratings as criteria is made evident by the apparently paradoxical finding of substantial and statistically significant relationships between (1) bombing data, which show no reliability, and (2) reliable superiors' ratings. This paradox is more apparent than real for it can be explained in terms of the contaminating effect of "unreliable" information shared in common by the raters.

Superiors' Ratings

The aircrews were rated as units by squadron or wing staff officers in terms of their performance in carrying out combat missions over Korea. The ratings were accomplished through the use of an eleven-item rating form. The eleven rating variables were:

1. *Skill as Technicians:* The degree of basic knowledge crew members have of their specialties as indicated by their performance or the degree of skill they exhibit in handling various equipments.
2. *Successful Completion of Missions:* The degree to which the crew reaches and bombs prescribed targets; including making necessary decisions in the absence of specific instructions and overcoming obstacles.
3. *Accuracy in Bombing Targets:* The accuracy with which targets are identified and bombed.
4. *Effectiveness of Crew Leadership:* The degree to which the aircraft commander organizes the crew to facilitate teamwork and cooperation among crew members.
5. *Consideration of Men on the Crew for One Another:* The extent to which crew members look out for the welfare of the crew as a whole, are liked by other men on the crew, and turn to one another as friends.
6. *Effectiveness in Working with Other Crews:* The degree to which the crew works as a part of a larger team and cooperates with other crews in carrying out a group effort.
7. *Effectiveness in Working with Superior Officers:* The degree to which the crew accepts orders or suggestions from superior officers and achieves objectives without conflict with superiors.
8. *Care of the Aircraft:* The degree to which the crew members insure proper maintenance of their aircraft and take personal interest in the plane and its equipment.
9. *Following SOP:* The degree to which the crew members carry out their functions in the prescribed manner.
10. *Military Bearing of Crew Members:* The degree to which members of the crew "conduct themselves in a military manner."
11. *Over-all Value to the Squadron (Wing):* The degree of over-all effectiveness of the crew as a part of combat unit.

Ratings on the items or variables were expressed as numerical values along a nine-point scale. Each point on the scale was defined, *nine* being the rating given to "undoubtedly the best crew in the squadron" and *one* being the rating given to "undoubtedly the worst crew in the squadron." All ratings were obtained in interviews with the raters.

* Based on a paper presented at the 1952 meeting of the American Psychological Association. The research was conducted under a contract between the Ohio State University Research Foundation and the Human Resources Research Laboratories (HRRL), Department of the Air Force. The data were collected in Japan and Okinawa by Lt. Col. Fred E. Holdrege and the senior author as representatives of HRRL. The authors wish to express appreciation to Mr. Luigi Petrullo, Chief, Personnel Research Directorate, HRRL, who initiated plans for the study of these crews. Interpretations and opinions expressed here are those of the authors and should not be regarded as having the endorsement of the Air Force.

The original plan was to secure a minimum of five independent ratings for each crew. This plan proved to be impractical due to the difficulty of locating raters who knew the crews sufficiently well to rate all variables. A total of 24 wing and squadron officers were utilized as raters in securing the ratings of the 94 crews. The mean number of ratings per crew actually obtained was 2.7 and the mean number of crews rated by each rater was 10.5.

An examination of the means and standard deviations of the ratings obtained from each of the 24 raters showed marked differences in their rating habits or bias. Before the ratings given a single crew by different raters were combined to form the final rating of the crew's performance, an adjustment was made to compensate for the observed rater bias. The adjustment was computed in such a manner that each rater's mean rating was approximately 50 and his standard deviation approximately 10.

The reliabilities of these converted ratings were estimated by application of a method developed by Horst (1) that is designed for the case in which varying numbers of raters are available. Intercorrelations of the eleven crew performance variables were also calculated. These intercorrelations along with the reliabilities, means, and standard deviations of the eleven variables are presented in Table 1. The reliability of these ratings ranges from .61 to .95 and meets standards which are generally acceptable for rating data.

The intercorrelations of the rating variables were factor analyzed and the final residual matrix and rotated factor loadings are presented in Table 2. Intercorrelations among the rating variables are relatively large. However, in addition to the large general factor, four additional factors are needed to account for the correlations. The five factors identified were:

I. General Reputation (Loadings on all variables); II. Objective Achievement (Loadings on 1, 2, 3, 4, 7 and 11); III. Technical Performance (Loadings on 1, 3, 9, 11); IV. Attention to Equipment and Procedures (Loadings on 8 and 9); and V. Social Performance (Loadings on 5 and 6).

Sociometric Nominations

On a sociometric nomination form, crew members were asked the following question: "If you could make up a crew from among the crew members in your squadron, whom would you choose for each crew position?" There were three general possibilities of reaction to the nomination question: (1) nomination of a

Table 1

The Reliability and Interrelationships of Eleven Variables of Aircrew Performance Rated by Superiors
Note: N = 83

Variable	Intercorrelation*											Mean	Standard Deviation
	1	2	3	4	5	6	7	8	9	10	11		
1 Technical Skill												49.6	8.21
2 Completion of Missions	84											51.3	7.64
3 Bombing Accuracy	74	68										47.7	9.65
4 Leadership	83	81	63									49.3	9.10
5 Consideration	62	66	56	70								50.3	7.30
6 W.W. Other Crews	69	69	63	74	73							49.6	7.72
7 W. W. Superiors	79	77	66	81	70	76						50.8	7.66
8 Care of Aircraft	66	64	47	63	61	67	56					50.1	6.93
9 Following SOP	80	73	62	78	59	68	75	66				50.2	7.74
10 Military Bearing	68	59	60	63	60	58	71	49	71			48.2	7.23
11 Over-all Value	85	78	73	82	67	78	78	61	82	62		47.4	10.25
Reliability	80	67	83	80	95	63	80	67	72	61	68		

* Decimal points have been omitted from these tables.

Table 2

The Final Residual Matrix and Rotated Factor Loadings Resulting from Factor Analysis of the Intercorrelations of the Eleven Crew Performance Variables

	Residual*											Rotated Factor Loadings*					
Variable	1	2	3	4	5	6	7	8	9	10	11	I	II	III	IV	V	h^2
1 Technical Skill												68	35	52	25	00	92
2 Completion of Missions	03											68	65	20	05	10	94
3 Bombing Accuracy	00	04										56	20	50	10	20	65
4 Leadership	−03	03	−08									69	30	40	30	15	84
5 Consideration	−01	−01	−03	00								67	20	10	20	50	79
6 W.W. Other Crews	00	01	01	00	−03							67	20	20	25	40	75
7 W.W. Superiors	00	00	−05	02	−02	03						67	30	40	10	30	80
8 Care of Aircraft	00	02	00	−06	03	04	−01					75	15	00	35	00	71
9 Following SOP	−06	06	−03	−04	−04	−02	03	−07				70	00	40	55	00	96
10 Military Bearing	04	01	04	−02	−01	−04	07	−06	06			65	10	25	15	20	56
11 Over-all Value	−01	06	01	−01	02	08	00	−03	−01	03		69	20	50	25	10	84

* Decimal points have been omitted from these tables.

crew member who was a member of the same crew as the nominator; (2) nomination of an individual from some other crew in the squadron; and (3) no responses. An individual usually indicated an "on-crew" choice by responding "same," "my own crew," or by writing the name of a fellow crew member in the blank provided. "Off-crew" choices were indicated by responding with the name of a man not on the individual's crew, and by such remarks as "Captain Smith's radio operator," or "anyone except the one we have."

The sociometric nomination data for each crew were used to compute an index of "on-crew" choices. The index is the ratio between the number of "on-crew" choices made and the total number of choices made. These index values ranged from .30 to 1.00 and were approximately normally distributed about a mean of .75.

In order to test the reliability of the index of "on-crew" choice, index values derived from random halves of each crew total nominations were correlated with one another. This correlation was .83 which, when extended by the Spearman-Brown formula, gave an estimated reliability of .91.

Bombing Error Criteria

The accuracy with which an aircrew is able to bomb combat targets may be considered as a near ultimate criterion of its effectiveness. Despite the high relevance of bombing data to the problem of evaluation of aircrew effectiveness, many conditions exist which detract from its utility. Chief among these are: (1) extremely variable conditions under which bombing must be accomplished; (2) severe limitations on the possibility of determining exactly where bombs are dropped in combat; and (3) limits upon the number of crews for whom the opportunity to perform the complete bombing operations exists (only the lead crew in formation bombing performs the complete bombing operation). Nevertheless, all available data were collected concerning the bombing accuracy in combat of the 94 FEAF crews. No combat bombing data whatsoever were available for 50 of the crews (no lead experience). Each of the remaining 44 crews had had one or more opportunities to lead formations on which it had been possible to secure photographs of where the bombs actually dropped. The bombing accuracy data consisted of *circular errors* for each of these bombing missions. These errors were expressed as the linear distance between the mean point of impact of all the bombs identified in the strike photograph and the assigned target. The number of circular error measures available for each of the 44 crews ranged from 1 to 8 with a mean of 3.16.

Inspection of the distribution of the bombing error data disclosed a markedly skewed distribution. A log transformation of these data yielded data with essentially normal distribution. The reliability of the transformed data as estimated, again by utilizing Horst's procedure, was not significantly different from zero.

Table 3 presents an analysis of the variance of the bombing data into between-crew differences and within-crew differences. It can be seen readily that there is approximately as much variance between the errors made by the same crew on different missions as there is between the mean performance of different crews. The difference between the mean bombing accuracy scores of these crews appears to be wholly unrelated to crew differences.

Relationships Among the Criteria

In order to complete the comparison of the three criteria, each of the eleven superior rating variables was correlated with both the sociometric index of "on-crew" choice and with the mean bombing error of each crew for whom we had bombing data. Table 4 presents these correlations.

Attention is called to the correlations between the bombing data and the superiors' ratings. It is quite apparent that the raters utilized the official bombing data as a source of information in making their ratings. Mean differences in the official bombing records of the crews serve to "contaminate" the raters' judgments on all rating variables, although the contamination is more marked for certain variables than for others. It is possible to extend the factor analysis of the eleven superior ratings to include the correlation with the bombing data by the following loadings on the five factors:

Table 3

An Analysis of the Variance of Transformed Bombing Accuracy Data

Source of Variance	Sum of Squares	Df	Variance	F ratio
Between Crews	208.56	33	6.320	1.34
Within Crews	451.82	96	4.706	
Total	660.38	129	5.119	

Table 4

Correlations Between Eleven Variables of Aircrew Performance Rated by Superiors and (1) the Index of "On-Crew" Choice and (2) Bombing Accuracy Data

Rating Variable	"On-Crew" Choice N	"On-Crew" Choice r	Bombing Data N	Bombing Data r	Bombing Data r corr.[1]
1 Technical Skill	90	.20	41	.58	.61
2 Completion of Missions	88	.10	41	.62	.70
3 Bombing Accuracy	80	.36**	41	.58	.58
4 Leadership	89	.13	41	.63	.67
5 Consideration	90	.10	41	.42	.48
6 W.W. Other Crews	90	.06	41	.47	.56
7 W.W. Superiors	90	.11	41	.57	.60
8 Care of Aircraft	90	.26*	41	.31	.38
9 Following SOP	90	.15	41	.40	.47
10 Military Bearing	90	.25*	41	.27	.30
11 Over-all Value	90	.18	41	.47	.54

[1] Corrected for restriction of the range of the rating variables (**2**).

* Significant at the .05 level.

** Significant at the .01 level.

I., General Reputation, .10; II. Objective Achievement, .80; III. Technical Performance, .40; IV. Attention to Aircraft and Procedure, .30; and V. Social Performance, .25

The major contamination is with Factor II which has its principal loading on the rating variable, Successful Completion of Missions. The original hunch responsible for the inclusion of this rating item was that it might identify variance in crew performance associated with low motivation and/or tendencies to abort missions. It appears, however, that information concerning the officially recorded performance of the crews determined much of this rating.

The sociometric index of "on-crew" choice does not appear to be related to any marked extent with the superiors' ratings. It may be noted that the larger of these small correlations are with rating variables which have high loadings on the Technical Performance and Attention to Equipment and Procedure factors but small loadings on the Social Performance factor. This fact suggests that nominations may well have been made on the basis of proficiency rather than just popularity. The

sociometric data may provide a second and relatively independent estimate of the performance of the crew. The correlation between the index of "on-crew" choice and the objective bombing accuracy data was found to be .33. This suggests that the sociometric choices were also influenced by the unreliable bombing information but to a lesser degree than most of the superiors' ratings.

Discussion and Conclusion

The finding of substantial and significant correlations between the objective bombing accuracy data, for which we had estimated a reliability of zero, and the eleven superior rating variables has general methodological implications for the development of criteria. An explanation of these correlations can be found in a possibility of a "contamination" of the judgments of superiors. The results of each bombing mission are widely publicized among the personnel of the wings and squadrons. In fact, it is standard procedure to hold a critique of the mission on the morning of the day following the mission. In addition, mission results, expressed in terms of bombing errors, are made part of official records which are maintained and used in determining which crews will be given an opportunity to lead further missions. It can be expected, therefore, that each rater could have estimated the official bombing accuracy records of each crew with a relatively high degree of accuracy. The fact that these records represented little other than chance crew achievement was, of course, unknown to the raters. This commonly shared, but unreliable, information concerning the performance of the various crews tended to produce both spuriously high reliability of the superior ratings and the spurious correlations with the unreliable bombing data.

A general question is raised concerning the dependability of rating data as criteria. In situations where objective achievement information is available, we may expect that raters will utilize such information in the process of forming the judgments they express in their ratings. If these achievement data reflect reliable performance, they will, of course, add to the dependability of the ratings. However, should the achievement information be basically unrelated to differences in the performance of the individuals or units being rated, this fact is likely to be overlooked when a test of the agreement between raters proves the rating to have "adequate reliability." Raters may agree in their knowledge of the achievement records but be in error about the meaningfulness of such records.

Received August 22, 1952.
Early publication.

References

1. Horst, P. A generalized expression for the reliability of measures. *Psychometrika*, 1949, **14**, 21–24.
2. Thorndike, Robert L. *Personnel selection*. New York: Wiley, 1949, p. 173.

Employment Prognosis of the Post-Poliomyelitic

Leonard V. Wendland

George Pepperdine College, Los Angeles, California

Within recent years the public has frequently been stimulated to think of its responsibility toward the physically handicapped person. There always have been those who were physically handicapped. To a large extent the crippled or deformed person either has been an object of special pity and attention, or he has been largely ignored. Certain crippling diseases, such as poliomyelitis, have rather consistently contributed their share of handicapped persons. Of those who contract poliomyelitis, approximately one-third retain some degree of residual physical involvement, while others are left with less noticable degrees of muscular weakness (**3**).

The individual's need to make a satisfactory psychosocial adjustment demands economic independence. This demands, on the part of the physically handicapped, as it does of the non-handicapped person, that he seek and retain employment. This employment must provide conditions and remuneration so that the employee will feel financially and emotionally secure. In this study we have surveyed the employment history of 151 post-poliomyelitic subjects, confining our analysis to the past 10-year period.

General Descriptive Data

Source of Subjects. In conjunction with a physical aftereffects survey carried out by the staff of the Orthopaedic Hospital, Los Angeles, California, in 1940, the Research Staff of the hospital carried on a study of the vocational, educational, and social status of the patients. The physical aftereffects study was made of randomly selected clinic patients, all of whom had had poliomyelitis and all of whom had been clinic patients of the Orthopaedic Hospital (**1**, p. ii). Clinic patients are by definition charity patients and must be under 21 years of age when treated. This age-limit becomes an important selective factor in the mean age of this sample of subjects. Such factors as sex, race, nationality, religion, and environmental background are accidental rather than selective.

Of the 794 patients available for study, those 16 years of age and older were contacted in the 1940 survey. Of the 794 patients, a total of 437, 203 males and 234 females, were interviewed by the Research Staff. The 437 patients represent all who were willing to cooperate in the study. The bulk of the data collected, in this 1940 study, has not been published, though a number of excerpts have been published (**4**, **5**).

In the present study every possible attempt was made to trace the present location of each of the 437 patients. Those patients residing in California, who were located and were willing, were used as subjects. Of the 437 patients, 40.3 per cent were located. Of those patients located, 85.8 per cent, or 151, were interviewed, 8.5 per cent were unwilling to cooperate and 5.7 per cent were deceased.

Age. The median age for the male subjects is 32.2 years, for females the median age is 34.7 years, and in terms of the total sample the median age is 33.9 years. An age-distribution such as we have would lead one to postulate that this group of subjects is at the present time approximately in what may be thought of as the "prime of life."

Age of Onset. Psychosocial adjustment to the disabling aftereffects of poliomyelitis may be directly or indirectly related to the age at which the subject contracted the disease. Of the group studied, 42.6 per cent of the males and 42.3 per cent of the females were under the age of 4 when they contracted the disease, and approximately 60 per cent of the males and 70 per cent of the females had the disease before their seventh birthday.

Residual Physical Involvement. The locale of physical involvement is important in at least two ways: (1) involvement in some areas is less disabling than others for the individual, and (2) certain types of involvement can be much more easily concealed than others. The degree of residual involvement is likewise

important in evaluating the psychological aftereffects. One general criticism by the subjects was that prospective employers tended to overemphasize their apparent physical limitations without giving them an opportunity to prove what they could do.

At the present time none of the subjects are under treatment for residual paralytic aftereffects. Each of the subjects was rated as to the apparency of the residual aftereffects on a five-point scale. A "0" apparency implies no apparent residual aftereffects, while a person rated as a "4" apparency is so seriously crippled that he is incapable of movements without the aid of major prosthesis and spends all of his time in a wheelchair or bed. In this group of subjects 62.3 per cent have either no apparent residual paralysis or only slight apparent involvement.

Employment History

The Degree of Physical Disability and Employment. It may be assumed that there is a rather direct relationship between the degree of physical involvement and employability. A comparison was made of males and females who are employed with those who are unemployed. Among the males, 89.7 per cent are employed and of the females 91.5 per cent are employed either within their own homes or otherwise. Approximately 57 per cent of the unemployed males have either no or only slight physical disability, while approximately 83 per cent of the unemployed females have either no or only slight physical involvement.

It is difficult to know how to interpret statements by seriously disabled individuals who say their physical condition does not interfere with their work. Among the males 57.3 per cent indicated that, in their opinion, they have definite physical barriers to certain types of employment. However, 78.0 per cent of this same group said that their physical disability does not interfere in their present employment. Two-thirds of the females feel that in spite of physical involvements their physical condition does not interfere with their present employment.

Description of Occupations Engaged in by the Group. The *Dicionary of Occupational Titles* (**2**) was used to classify all those employed. Table 1 gives a summary of males and females

Table 1

Summary of Occupations

Occupational Description	Male		Female	
	N	%	N	%
Professional	20	29.4	6	7.2
Semi-professional	3	4.5	3	3.6
Sales	8	11.8	2	2.4
Clerical	8	11.8	16	19.4
Service	1	1.5	44	53.0
Agricultural services	2	2.8	0	0.0
Skilled	16	23.5	3	3.6
Semi-skilled	3	4.5	1	1.2
Unskilled	0	0.0	2	2.4
Unemployed	7	10.2	6	7.2
Total	68	100.0	83	100.0

in terms of this standard description of occupations. It is worthy of note that 29.4 per cent of the males are employed in what is rated as professional work. When the number of male subjects engaged in work classed as professional is compared with the general population we note that while 29.4 per cent of this group are so employed, only 4.5 per cent of the general Los Angeles population are employed in work of this nature (**6**, p. 190).

Type of Employer. In "business for themselves" accounts for 26.4 per cent of the male subjects. The next largest percentage in this group is absorbed by those in the small business concern, while 19.1 per cent of the group are employed in some large public corporation. Next to domestic work as housewives, the females in this study are primarily attracted to public service employment. Of the females, 9.6 per cent, however, are in business for themselves.

Weekly Income in Relation to Degree of Disability. The median weekly income for males is approximately $74.64, while for female subjects it is approximately $59.71. About 25 per cent of the male subjects earn more than $100.00 per week, while a small number have weekly incomes in excess of $200.00.

Part A of Table 2 gives the mean weekly incomes for males and Part A of Table 3 gives similar data for females, in terms of residual physical involvement. Parts B of Tables 2 and 3 make comparisons of groups of subjects with varying degrees of residual paralysis and mean incomes, giving the statistical level of

confidence for each comparison. It might be assumed that individuals with little or no residual paralysis should have a greater mean income than those more seriously handicapped. Part B of Table 2 indicates that this assumption is not borne out. Part B of Table 3 does, however, confirm this contention. Some of the relationships in Parts B in Tables 2 and 3 have little statistical significance, they are given, however, to indicate the spread found in the relationships.

These earnings reflect a significant rise in weekly income when compared with the earnings of this group prior to World War II (**5**). Since 29.4 per cent of the males have professional employment, the median income of these male subjects has been compared with the median income of professional employees of the general Los Angeles population. The annual median income for subjects employed in professional positions is about $5,124.50, whereas the median income for comparable employees of the general Los Angeles population is $3,972.00 (**6**, p. 294). By interpolation we find that the median income for the subjects under study is approximately at the seventy-first centile level when compared with the median income of the general Los Angeles population.

Table 2

Mean Weekly Income of Employed Males in Terms of Handicap—with Level of Confidence of Relationships

A

Degree of Apparency	Mean Income	Total N	Total %
4	50.00	1	1.6
3	96.46	10	16.1
2	126.56	14	22.6
1	79.44	27	43.6
0	94.51	10	16.1
		62	100.0

B

Inter-Group Relationship	Statistical Level of Confidence in Rank Order
0–1	.30
1–4	.30
2–3	.40
2–4	.40
3–4	.50
0–4	.50
1–2	.05
1–3	.30
0–2	.40
0–3	.90

Table 3

Mean Weekly Income of Employed Females in Terms of Handicap—with Level of Confidence of Relationships

A

Degree of Apparency	Mean Income	Total N	Total %
4	22.57	3	8.6
3	49.74	8	22.9
2	93.50	7	20.0
1	60.13	13	37.1
0	62.50	4	11.4
		35	100.0

B

Inter-Group Relationship	Statistical Level of Confidence in Rank Order
1–4	.01
0–4	.05
3–4	.05
1–3	.10
2–3	.10
2–4	.10
0–3	.30
0–1	.80
1–2	.10
0–2	.40

About 24 per cent of the male subjects are in some form of skilled employment. The median annual income for these males is approximately $3,874.50, whereas the median income for comparable employees of the general Los Angeles population is $2,746.00 (**6**, p. 294). By interpolation of data provided, we find that the median income of the subjects studied is probably in the vicinity of the eighty-third centile level when compared with the median income of the general Los Angeles population.

Since 19.4 per cent of the females hold some type of clerical position, a comparison was made between the median annual income of these subjects and the median annual income of like employees of the general Los Angeles population. The median annual income for these subjects is approximately $2,999.50, as compared with the $1,728.00 median annual

income for the general Los Angeles population (**6**, p. 295). By means of interpolation it may be ascertained that the median annual income for the subjects employed in clerical positions is at approximately the ninety-fourth centile level when compared with the median annual income of the general Los Angeles population.

The comparisons made above indicate that the annual income for males employed in professional and skilled positions, and for the females employed in clerical positions, is in each instance above the median annual income of the general Los Angeles population in similar employment. In most other comparisons it was found that the income of the subjects of this study compared favorably with the general Los Angeles population.

Relation of Educational Level Reached and Weekly Income. In making a study of the educational level reached as it related to the mean weekly income, it was found that, generally, the higher the educational level reached the higher the income. Table 4 indicates that for males, with the increase of education there is an increase of the mean weekly income in nearly every instance. Table 5 shows that those females with the most education are to be found in positions in which the highest salaries are earned. Parts B of Tables 4 and 5 give the statistical significance for some of the relationships between educational level reached and mean weekly income.

Table 4

The Significance of Mean Income of Male Subjects in Relation to Education

A

Group Code	Educational Level Reached	Total N	Weekly Mean Income
1	Professional degree	7	$162.15
2	Graduate study	4	70.00
3	College graduate	6	96.54
4	1–3 years additional study	19	97.22
5	0–1 year additional study	14	86.29
6	High school graduate	6	77.51
7	High school non-graduate	5	57.98
		61	

B

Inter-Group Relationship	Statistical Level of Confidence in Rank Order
3–7	.05
1–4	.05
1–5	.10
1–6	.10
1–7	.10
4–7	.10

Table 5

The Significance of Mean Income of Female Subjects Employed Outside of Home in Relation to Education

A

Group Code	Educational Level Reached	Total N	Weekly Mean Income
1	Professional degree	1	$269.00
2	Graduate study	3	84.58
3	College graduate	2	58.75
4	1–3 years additional study	10	55.92
5	0–1 year additional study	15	50.05
6	High school graduate	1	41.75
7	High school non-graduate	1	48.00
8	Grade school only	1	53.20
		34	

B

Inter-Group Relationship	Statistical Level of Confidence in Rank Order
1–4	.01
1–5	.01
2–5	.01
1–2	.05
1–3	.05
2–4	.05

The Role of Prosthesis and Employability. Approximately one-half of the 61 employed males do not wear any type of prosthesis. Another 10 per cent of the males wear only some form of orthopaedic shoe; that is, a shoe with a lift built into it or some other form of correction.

About 42 per cent of the females who are employed outside of the home do not require any form of prosthesis, while 8.3 per cent of this group require leg braces and are confined to a wheel chair. Approximately 55 per cent of the females require some form of prosthesis as compared with 35 per cent of the males.

Of those males and females who are at present unemployed, 71.4 per cent do not require any form of prosthesis. Accordingly we may assume that those males and females who are unemployed are not primarily so, due to physical involvement, but that there are other more significant factors.

Opportunities for Advancement in Present Employment. Employee morale is heightened when one feels that his employment offers opportunities for advancement. About 57 per cent of the males and 33 per cent of the females feel that they have employment which offers opportunities for advancement.

Job Stability. Another indication of occupational adjustment may be inferred from job stability. It is assumed that a lesser number of employment shifts made within the last 10-year period is indicative of a better occupational adjustment than is frequent employment shifting.

About 25 per cent of the male subjects have been employed by the same employer for at least the last 10-year period; and 57.1 per cent of the male subjects have had a maximum of 3 different employers during the last 10 years.

About 78 per cent of the females have had no more than 3 different employers during the last 10-year period. About 18 per cent of the females have been employed in the same establishment during the last 10 years.

An analysis of data was made regarding the relationship of job stability and residual physical involvement. The indications are that those subjects with no evident residual paralysis and those most severely handicapped had a greater tendency to shift jobs than did those subjects with moderate to moderately severe residual physical involvement.

Subjects Who Had Been Employed in "War Work." The last 10-year period includes a number of years in which the nation's economy was organized in terms of the materials required by the war effort. During the war years labor was at a premium. Individuals who, previous to the war, had had difficulty finding employment now found that their services were being solicited. About 62 per cent of the males and 27.5 per cent of the females have been employed in "war work." All of the subjects who took "war work" had had previous employment or were self-employed.

In this study 10 males were rated as having no apparent residual involvement. Of these males, half saw active service during the recent war, in some branch of the armed forces.

Summary

In this paper we have noted the following relevant data obtained from a study of the past 10-year employment history of 151 post-poliomyelitics.

1. About 90 per cent of the males and 43 per cent of the females are employed outside of the home.

2. About 26 per cent of the males and 10 per cent of the females are owners of business establishments.

3. The median weekly income for males is approximately $73.64, while that for females is $59.71.

4. There is, in some cases, a significant relationship between education and income, with the higher incomes being earned by those with the most education.

5. About 55 per cent of the females and 35 per cent of the males require some form of prosthesis. Neither serious physical involvement nor the need for prosthesis seems to be a major factor for those subjects who are at present unemployed.

6. About 25 per cent of the males and 18 per cent of the females have been employed in the same employment for the last 10 years.

7. It was indicated that 62 per cent of the males and 28 per cent of the females have had employment related to the nation's war effort. Half of the males with no evident residual involvement served in the armed forces.

8. The data would indicate that the employment prognosis of the post-poliomyelitic compares favorably with that of the non-handicapped population of the Los Angeles area.

Received September 28, 1951.

References

1. Carroll, R. L. Aftereffects survey of infantile paralysis at Orthopaedic Hospital, Los Angeles. Unpublished, Orthopaedic Hospital Library, Los Angeles, California, 1941.
2. *Dictionary of occupational titles.* Volume I, Definition of titles. (Second edition, March 1949.) Government Printing Office, Washington, D. C., 1949.
3. *Facts and figures about infantile paralysis.* New York: National Foundation for Infantile Paralysis, Publication No. 59 (second edition), 1946.
4. Lowman, C. L., and Seidenfeld, M. A. A preliminary report of the psychosocial effects of poliomyelitis. *J. consult. Psychol.*, 1947, **11**, 30–37.
5. Seidenfeld, M. A. Psychological elements in work interference from physical disability. *J. consult. Psychol.*, 1947, **11**, 326–333.
6. *Statistical abstracts of the United States.* U. S. Bureau of Census. (Seventh edition.) Washington, D. C., 1949.

A Job Preference Survey for Industrial Applicants*

W. F. Long**

Purdue University

In a comprehensive discussion of the nature of interests, Super (**10**) states that interests have probably received more attention from vocational psychologists during the past generation than any other single type of human characteristic, including intelligence, aptitudes, and personality traits.

In spite of this considerable expenditure of effort, few, if any, of the large number of interest assessment devices which are currently available, are entirely satisfactory for use in business and industry as an aid in the selection and placement of unskilled and semi-skilled workers. All of those in general circulation have one or more limiting characteristics for this purpose. Almost all are intended to focus interests into occupational areas, nearly always at the professional or semi-professional level. The few inventories and information tests which have been constructed for use at lower skill levels are too specific for general use. Many of the inventories or questionnaires include items which require expressions of interests concerning activities, jobs, hobbies, and the like, about which the respondent could have but limited first-hand information. Perhaps the severest limitation to the utilization of published interest inventories for selection and placement is their susceptibility to falsification by respondents who attempt to match their expressed interests to those they believe necessary for a desired job.

This study is an attempt to develop a *Job Preference Survey*[1] which is relatively free from the mentioned limiting characteristics. The *Survey* is intended to sample job activity preferences at a non-professional level. The interests expressed are reflected as relative preferences for antagonistic pairs of job activity types, e.g., routine versus creative work, rather than for occupational areas. The majority of items describe unitary tasks with which most respondents from a population of business and industrial applicants would be familiar. The form and content of the items were chosen in an attempt to enhance validity by simplicity of presentation while at the same time minimizing the possibility of falsification by job applicants.

Construction of Job Preference Survey

Selection of Job Characteristics for Study. In selecting those job characteristics to be studied as components of work interests, reference was made to psychological literature and to several texts and treatises on job analysis. Although it was found that fundamental job characteristics have not been clearly identified, eight antagonistic pairs of job activity types were selected for study as possible independent components of work interest.[2] These components are: 1. Routine—Creative; 2. Indoor—Outdoor; 3. Repetitive—Varied; 4. Responsible—Non-responsible; 5. Hazardous—Non-hazardous; 6. Sedentary—Bodily Active; 7. Isolative—Gregarious; and 8. Precise—Approximate.

Construction of Items. The paired-statement item form was used because of the nature of the problem at hand which makes it necessary to require a choice to be made between two activities representative of the extremes of a hypothetical continuum. The couplet form is also believed to be less confusing to the low

* This article is based upon a thesis submitted in partial fulfillment for the degree of Doctor of Philosophy, done under the direction of Professor C. H. Lawshe. Complete data for the study can be found in the author's thesis on file in the Purdue University library.

** The author is now a major in the Air Force of the United States assigned to duty as Executive Officer, Personnel Research Laboratory, Human Resources Research Center, Air Training Command, Lackland Air Force Base, San Antonio, Texas.

[1] Copyright by the Purdue Research Foundation and distributed by the Occupational Research Center, Purdue University, Lafayette, Indiana.

[2] After this project had developed so that any modification of the components under study would be impractical, reports by Kelley (**4**) and Vernon (**11**) were located which suggested two additional components, viz., Things or Mechanisms—People and Verbal Activity—Physical Activity. It is intended to include these in an extension of the current study.

ability job applicant and therefore less likely to introduce extraneous variance. The pairs were formed so that the prestige or social acceptability level and the performance skill level were judged by the author to be similar. It obviously was also necessary to make each pair of statements suggest a choice within one component only. Forty items were constructed for each of the eight components.

The instructions given to the subject, which include a sample item from the Sedentary—Bodily Active component, follow:

> In the Job Preference Survey you will find job activities listed in pairs. Choose the one of each pair which you would like best if you could perform both activities equally well, and would receive the same pay. Make a check mark (√) after the job activity in each pair which you would like best. Mark one and only one in each pair.
>
> Example:
>
> Keep records of materials in stock____( )
> Deliver materials to machine operators____(√)
>
> For the example pair of job activities, you would place a check mark as is shown on the line to the right of "Deliver materials to machine operators" if you would like best to perform that one of the two job activities.
>
> Work quickly. Do not spend too much time thinking about one pair of activities. You probably will be able to complete the Survey in just a few minutes.

Example items from each of the other seven components follow:

1.	R–C	Pack glassware for shipment. Design decorative patterns for glassware.
2.	I–O	Operate a large batch mixer in a foods plant. Operate a large mixer making ready-mix concrete.
3.	R–V	Pack shirts in a factory shipping department. Wrap articles in the delivery department of a retail store.
4.	R–NR	Supervise a crew of carpenters. Work as an expert carpenter.
5.	H–NH	Pour melted metals from overhead ladles into molds. Prepare molds for making heavy castings.
7.	I–G	Look up statistical data in a reference library. Operate the loan desk of a library.
8.	P–A	Lay individual tiles for a shower room wall. Lay concrete blocks for a building partition.

Construction of Experimental Forms. In order to keep administration time within reasonable limits, four forms were constructed, each including 80 items covering two of the components. Within each form the components were paired with the intent to minimize similarity of item content and the items were arranged in a randomized fashion to reduce the effect of repetition of component and continuum extreme.

Administration and Scoring of Experimental Forms. By arrangements made through the offices of the Occupational Research Center, Purdue University, from 166 to 174 copies of each of the experimental forms were completed by job applicants from 20 companies. Each of the experimental forms was scored for both components in terms of the number of preferences indicated for one of the extremes of each component. For example, if a subject indicated a preference for indoor over outdoor job activities in 28 of the forty items, his score in that component would be 28.

Item Analysis and Revision of Experimental Forms. Based upon upper-lower half splits, a measure of internal consistency was determined for each item in a component as well as a measure of relationship of response for each item of a given component with the total score on the other component in the same form. These relationships were quantified using D-Values (**6**). The best 25 items of the original 40 in each component were selected for inclusion in a single revised form, Form **X**. Those items were retained which had the highest internal consistency values and at the same time the lowest relationship with the other component in the first experimental form. No item was retained which had an internal consistency D-Value of less than .45 or a D-Value based on the "other" criterion which was larger than its internal consistency D-Value. In order to reduce the effect of item position upon subject response in Form **X**, the items were arranged in a random manner. The items were also arranged so that either extreme of the components was listed first an approximately equal number of times.

Administration and Scoring of Form X. A total of 260 usable completed copies of the *Survey* were returned by twelve additional companies which had agreed to participate in the project.

The completed *Survey* forms were again scored in terms of the number of preferences indicated for one of the extremes of each component. Thus eight scores were secured for each subject.

Double Cross-Validation Design for Further Revision. The 260 completed copies of Form X were divided between two equal groups, A and B, in order to permit use of the double cross-validation method as described by Katzell (**3**). This design provides for scoring of Group A responses using a key based upon Group B responses and vice versa. A final key is developed using the item analysis results from both samples.

The basic principle underlying the double cross-validation method is the enhancing of the reliability of findings through the replication of experiments. Greater confidence can be placed in congruent results from two or more independent samples than in the results of a single experiment, where the total N is the same in the two experimental designs.

When applying the method to an internal consistency analysis, two estimates of reliability are secured, each based on a larger sample than probably would have been available using the conventional method. These two reliability estimates are probably underestimates of the reliability of the final key because if the same number of items are included in all three keys, those in the final key will be more stable than those in the two sample keys.

Item Analysis of Form X. Item counts were completed separately for Groups A and B. Based upon upper-lower half splits, a measure of internal consistency was determined for each component as well as a measure of relationship of response for each item of a given component with the total score on the other seven components. These relationships were quantified into *omega* values (**7**) which in turn are easily translated into probability values, i.e., the confidence level at which the differences between upper and lower halves may be said to be different from zero.

Development of Scoring Keys. Using the item analysis results from Group A, the best 20 items from the total of 25 in each component were retained to make up a key to be used to re-score the forms from Group B and vice versa. The decision to retain 20 items was arbitrary, being a compromise considering reliability and testing time. Items were retained so as to maximize internal consistency while at the same time minimizing relationships with other components. From the data for Group A it was possible to select 146 of the 160 items which discriminated between the upper and lower halves at less than the one per cent confidence level on an internal consistency basis. The remaining items discriminated at less than the 12 per cent level. From the data for Group B it was possible to select 155 of the 160 items which discriminated at the one per cent level. The remaining items discriminated at less than the 12 per cent level.

Similarity of Groups and Keys. None of the differences between the means and standard deviations for the two samples is significant at the five per cent level of confidence. Thus, it can be said that neither the two forms nor the two sample groups differ appreciably.

Intercorrelation of Scales. The intercorrelations between the scales for Groups A and B are presented in Table 1. The development of Scale 9 listed in the table will be discussed subsequently. The values given are product-moment coefficients, which were considered to be more suitable than tetrachoric coefficients in view of the small size of the samples (**2**). Although the size of the correlations varies somewhat between samples, the general relationships are consistent.

The intercorrelations among the Routine—Creative, Repetitive—Varied, and Responsible—Non-responsible scales indicated that revision would be desirable. Accordingly, a composite scale was constructed using items from the Routine—Creative and Repetitive—Varied groups. Items were selected in an attempt to make the relationship with the Responsible—Non-responsible scale as small as possible. Using the composite scale keys, the *Survey* forms were scored in the same manner as previously. From Table 1 it can be seen that the composite scale, identified as Routine—Varied, is also substantially correlated with the Responsible—Non-responsible scale. It would appear that the responsible and non-responsible activities described in the items differed substantially in terms of the variety of activities involved as well as in terms of degree of responsibility. It therefore seemed

Table 1

Reliabilities and Intercorrelations of Scales in Form A*

Scale	Group	Scale 1	2	3	4	5	6	7	8	9
1. Precise	A	81	−50	16	32	−05	23	47	−30	−37
—Approximate	B	86	−50	16	35	−05	38	39	−46	−44
2. Routine	A		83	−19	−11	37	−67	−27	67	
—Creative	B		85	−50	−17	15	−83	−11	65	
3. Hazardous	A			78	00	−07	28	01	−16	−17
—Non-hazardous	B			82	−03	−04	48	−14	−38	−47
4. Sedentary	A				66	20	02	50	15	−03
—Bodily Active	B				68	39	00	52	06	−04
5. Isolative	A					70	−46	18	40	40
—Gregarious	B					79	−27	19	23	16
6. Responsible	A						79	12	−39	−70
—Non-responsible	B						81	07	−64	−77
7. Indoor	A							75	−04	−16
—Outdoor	B							74	−06	−05
8. Repetitive	A								78	
—Varied	B								76	
9. Routine	A									82
—Varied	B									80

* Decimals omitted. Negative coefficients indicate positive relationships with reverse extreme of scale continuum.

advisable to drop the Responsible—Non-responsible and to retain the composite Routine—Varied scale leaving six scales in the *Survey*. Only two of the intercorrelations for both groups among the six scales retained are as high as .50. Most of the remainder are much lower.

Reliability of Scores. The reliability estimates for the keys used to score Groups A and B are given in Table 1. These estimates were determined using the Kuder-Richardson Case III, Equation 20 (**5**). It is to be remembered that in the double cross-validation design these obtained estimates probably are underestimates of the reliabilities of the final keys and that the larger of the two is likely to be the more accurate estimate. Under these assumptions, three of the scales would have reliability coefficients which are at least larger than .80, two well above .70, and one nearly.70. It is believed that these reliabilities can be considered to be satisfactory inasmuch as they are based upon only 20 items. It is also believed that test-retest reliability coefficients would be higher than the internal consistency ones here reported.

Construction of Composite Key. The selection of items to be included in the final key using a double cross-validation design is based upon consideration of the discriminative power of the items in both sample groups. In this study the significance of the differences between preferences for statements in a given scale made by the upper and lower halves divided on the basis of total score were translated into probability values, i.e., the confidence level at which the difference may be said to be different from zero. The composite probability value for each item was determined by calculating the product of the two probability values.[3] In the selection

[3] Lindquist (**8**) suggests that the use of chi-square is a more appropriate means for determining composite probability. However, since the same rank order of items is secured by either method, the less complex one was used. This method results in the selection of the same items, but according to Lindquist yields inflated probability estimates.

of items, relationship to other scales was considered as well as internal consistency. The five least satisfactory items in each scale were designated for elimination leaving 20 items in each scale. No item was retained that had a composite internal consistency probability value greater than .0004.

Considerations in Possible Application of the Job Preference Survey. With an instrument such as this *Survey*, interest profiles can be truly contrasted, with allowance for differences in scale means, as cannot be done with the usual preference type of inventory wherein an expressed interest for one activity usually demands an omitted preference for a different type of activity, thus exaggerating profile differences. Furthermore, in the *Survey* the preferences required are within one category, not between categories.

It is not essential to have population norms or distributions in order to use the *Survey* results for placement purposes since the relative strength of interests is meaningful. Thus, individual profiles can be compared to synthetic standards determined by job analysis designed for that purpose. It is true, however, that an empirical profile of successful workers (in the broadest sense of the term) for each job would be most useful. Both the synthetic and empirical profiles probably would be most meaningful in terms of the range of scores made by a given proportion of the population, viz., the middle 50 per cent.

Summary

The importance of interests to job success and satisfaction has been recognized for some time. However, probably none of the large number of interest assessment devices which are currently available are entirely satisfactory for use in business and industry as an aid in the placement of unskilled and semi-skilled workers. This study was undertaken to develop a *Job Preference Survey* to meet this apparent need.

The final form of the *Survey* includes 120 paired-statement items, descriptive of six essentially independent components of work interest which are: 1. Routine—Varied; 2. Indoor—Outdoor; 3. Hazardous—Non-hazardous; 4. Sedentary—Bodily Active; 5. Isolative—Gregarious; and 6. Precise—Approximate.

Only one of the intercorrelations among the six scales is as high as .50 while most of the others are much lower. Three of the scales were found to have internal consistency reliability coefficients in excess of .80, two well above .70, and one nearly .70. These coefficients are conservative estimates of the internal consistency of the scales and it is believed that test-retest reliability coefficients will be found to be appreciably higher.

It was suggested that an instrument such as the *Survey* can be used profitably in job placement with reference to comparison of individual profiles with an empirical standard, or to comparison of profiles with synthetic standards based on job demands determined by job analysis. These comparisons would probably be most meaningful in terms of the range of scores made by a given proportion of the population, viz., the middle 50 per cent.

Received October 5, 1951.

References

1. Fryer, D. *Measurement of interests.* New York: Henry Holt, 1931.
2. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill Book Co., 1942, 244.
3. Katzell, R. A. Cross-validation of item analyses. *Educ. psychol. Measmt*, 1951, **11**, 16–22.
4. Kelley, T. L. Report on an activity preference test for the classification of service personnel. *OSRD Report No. 4484.* Cambridge: Harvard Univ., 1944.
5. Kuder, G. F., and Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, **2**, 151–160.
6. Lawshe, C. H. A nomograph for estimating the validity of test items. *J. appl. Psychol.*, 1942, **26**, 846–849.
7. Lawshe, C. H., and Baker, P. C. Three aids in the evaluation of the significance of the difference between percentages. *Educ. psychol. Measmt*, 1950, **10**, 263–270.
8. Lindquist, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin Co., 1940, 46.
9. Strong, E. K., Jr. *Vocational interests of men and women.* Palo Alto: Stanford Univ. Press, 1943.
10. Super, D. E. *Appraising vocational fitness.* New York: Harper & Bros., 1949, 376.
11. Vernon, P. E. Classifying high-grade occupational interests. *J. abnorm. soc. Psychol.*, 1949, **44**, 85–96.

Development of a Home Economics Interest Inventory[1]

Hildegarde Johnson

Iowa State College

The study reported here is the first of a series planned to develop a special vocational interest inventory in the field of home economics.

Such a special vocational interest inventory differs from the general inventories which are widely used in the guidance of students in that it is designed to differentiate among occupations in a limited field. General vocational interest inventories sample a wide variety of interests in an attempt to include some interests of persons in every occupation as an aid to classifying a student's interests into broad types, such as persuasive or mechanical, or to identifying a student's interests with those of one of a large number of specific occupations. In contrast, the items in a special vocational interest inventory are a sample of the interests which bear upon occupations within one field and the inventory may be used to compare the vocational interests of a student who has selected this field with the interest patterns of the specialists in it.

The need for a special inventory in the field of home economics has been recognized for some time. Attempts which have been made to develop keys for occupations in home economics on the Strong and Kuder tests have resulted in differentiation of some occupational groups from women-in-general. However, the percentage of overlapping between the distributions is so great that the usefulness of the keys for individual guidance may be questioned.

Basic Assumptions

The first assumption made in this study is that interests of persons employed in a given occupation form a pattern. This assumption may be stated as a theory which is consistent with research in the area of interests: if a number of interests and aversions which are characteristic of persons in a specific occupational group can be identified and their average intensity determined, it is possible to delineate a pattern in which the component parts are the interests and aversions and the average degree of intensity represents the dimensions of each part. An occupational interest pattern may be defined, then, as a configuration of intensities of interests and of aversions which are common to persons in an occupational group but not common to persons-in-general. The ease of determining such a pattern for an occupational group depends upon the homogeneity of persons in the group and the extent to which the true interests and aversions which belong in the pattern occur in the sample of interests which are included in the inventory.

The second assumption is that if a student has an interest pattern which is similar to the interest pattern of those in an occupational group, he will find greater satisfaction in that occupation than he would find in one in which the persons employed have a different interest pattern. It is reasonable to suppose that the average person in an occupational group is achieving vocational satisfaction. This satisfaction probably results from a reaction of interest rather than of aversion or indifference to the activities which are performed on the job and to the psychological stimuli which are concomitant with the job. If this is true, another person who has the same interests may be expected to respond in the same way to the same activities and psychological stimuli and he would, therefore, find satisfaction in the same job. Research indicates that vocational interests change relatively little during the college years and after graduation; hence, it is reasonable to suppose that if a college student has the same interests as professional persons in a specific occupation, he would respond in the same way, not only at the time his interests are measured as a freshman in college, but a few years later when he is ready to accept a position. Assuming suitable abil-

[1] This paper is largely a summary of a Ph.D. dissertation on file in the Iowa State College Library. The writer expresses sincere appreciation to Dr. Hester Chadderdon who directed the study.

ity, a student will find vocational satisfaction if he selects the occupation for which the interest pattern is most like his own.

Development of the Trial Form

Items in a special vocational interest inventory should be a sample of the interests and aversions which are components of the true interest pattern of occupations within one field. These interests have not been identified and cannot, therefore, be sampled by the usual methods of sampling. However, they are interests and aversions associated with a finite number of jobs within a vocational field. Therefore, the assumption seems reasonable that an adequate analysis of occupations in the field would identify many of these interests and would therefore be the best basis for the selection of items. Analyses of the occupations in home economics were made by reading vocational guidance literature and by interviewing 26 persons who had had vocational experiences. Additional items for the trial form of the home economics interest inventory were secured from data collected in a previous study of interests of home economics students[2] and from general vocational interest inventories in current use.

The 448 items in the trial form of the inventory were grouped into three sections. The first group of items were *activities* to which reactions were to be indicated on this five-point scale: (1) like very much (2) like (3) indifferent, or do not know (4) dislike (5) greatly dislike. An example of an activity in this section is item 51, "supervise preparation and serving of food."

The second group of items was *job characteristics and environmental factors* to which reactions were also to be indicated on a five-point scale: (1) highly desirable characteristic of a job (2) desirable (3) not important (4) undesirable (5) highly undesirable. An example in this section is item 345, "work in which it is possible to see definite change in people."

The third section was made up of miscellaneous items combined into series to be ranked in order of preference. Items 359 to 362, for example, are: "work with preschool children," "work with grade school children," "work with adolescents," and "work with adults."

[2] Nelson, Esther. *A device to determine job characteristics influencing college home economics students in their choice of a major.* Unpublished M.S. thesis. Ames, Iowa: Iowa State College Library, 1947.

Collection of Data

Fourteen home economics occupations were selected for study, criterion groups being made up of a sample or a census of persons employed in each of these occupations. To secure these groups a total of 1,884 inventories was mailed. From the 1,799 inventories which presumedly reached their destination, 1,175 or 65 per cent responses were received. Since the returns for four of the occupational groups proved to be too few for satisfactory study, the analysis was limited to ten occupational groups. To balance the number of returns in the ten groups and facilitate the computations, one hundred returns were selected at random from each of eight groups and all of the inventories were used in the two remaining groups.

Criterion groups are described briefly in Table 1 in which are recorded the number of persons, their median age and the median years of experience they had had in the occupations in which they were employed at the time they responded to the inventory.

Analysis of Data

Responses to the trial form of the inventory of each of the ten groups were tabulated separately, using an IBM test scoring machine. Chi-square was the statistic selected to determine which items differentiated among occupational groups. When this statistic is computed it is not necessary to assume that responses are normally distributed or that the responses used form a true scale. Furthermore, chi-square was eqully appropriate for items in the first two sections in which responses were indicated on a five-point scale and for items in the last section which were ranked in order of preference.

A technique for analyzing the data was desired whereby not only the items which differentiate among occupational groups would be identified but by which items could be ranked for each occupational group according to the extent to which they differentiate a

given group from other home economists. The contribution of each occupational group to the chi-square value for each item was used as this basis for ranking items and to determine for which occupational groups the item was discriminating. The distribution of responses which was originally indicated on a five-point scale in the first two sections of the inventory was broken into two groups to facilitate computation of the components of chi-square. Items in which the distribution of the combined samples was skewed toward the unfavorable responses were divided by combining the neutral responses with the unfavorable responses. Items skewed toward the favorable responses were divided by combining the neutral responses with the two favorable responses.

The formula selected to compute chi-square and the individual contributions of each occupational group to the chi-square value is one that has not been commonly employed. It is:

$$\chi^2 = \sum_{i=1}^{n} \frac{(a_i N_2 - a_i' N_1)^2}{N_1 N_2 (a_i + a_i')}$$

n = total number of occupational groups,
i = occupational groups,
a_i = number of responses of "like" for each occupational group,
a'_i = number of responses of "dislike" for each occupational group,
N_1 = sum of responses of "like,"
N_2 = sum of responses of "dislike."

For each item there are ten component parts of chi-square, one for each occupational group. Table 2 is an example of the results of analyses of items.

Ninety-two per cent of the 448 items in the inventory were significant at or beyond the 5 per cent level and 89 per cent were significant at or beyond the 1 per cent level. Therefore, most of the items included in the trial form are of value in differentiating among these ten occupational groups. This would seem to indicate that the techniques used to find items were highly satisfactory.

Items were selected for ten scoring keys by determining which items most successfully differentiate each occupational group from other home economists, the size of the numerical contributions of this group to chi-square values forming the most important basis for selection. An attempt was also made to balance the number of items in the various keys and to balance the number of items in each to which persons responded mover favorably and to which they responded less favorably than did other home economists.

In the first study weights of 4, 3, 2, 1, and 0 were tentatively assigned to the levels of response to items selected for the scoring keys. Further experimentation with methods of weighting is planned after which weights will be used which result in successful differentiation among groups with reasonable expenditure of time in scoring.

The Home Economics Interest Inventory and Strong's *Vocational Interest Blank for Women* were compared with respect to the degree of differentiation achieved between home economics teachers and a base group. The distributions of scores on the Home Economics Teacher key of the *Home Economics Interest Inventory*

Table 1

Number of Persons, Median Age, and Median Years of Experience of Criterion Groups

Occupation	No. of Persons	Median Age	Median Years Experience
Extension home economist	100	35.0	5.0
Food product promotion	100	36.2	4.8
Food service director	100	36.2	5.7
Group work with young children	100	28.2	4.0
Home economist in social welfare	100	37.0	5.2
Home service director	100	31.3	4.2
Hospital dietitian	100	38.5	12.5
Journalist or radio home economist	94	35.2	5.3
Restaurant and tea room manager	69	33.5	7.5
Teacher (high school)	100	39.8	5.3

Table 2

Chi-square Value and Components of Chi-square for Item 32

Occupational Groups	Frequencies		Components of Chi-square
	Upper Level	Lower Level	
Extension home economist	35	65	1.0332
Food product promotion	44	56	.6737
Hospital dietitian	28	72	5.9803
Teacher	44	56	.6737
Home service director	47	53	2.0542
Food service director	24	76	10.6408
Group work with young children	33	67	2.0299
Journalist or radio home economist	64	30	30.9447
Home economist in social welfare	51	49	5.0616
Restaurant and tea room manager	15	54	9.5668
Total	385	578	68.6589

of 100 home economics teachers and 100 home economists-in-general were estimated to overlap[3] 23 per cent, and no women in the home economist-in-general group score equal to or higher than the mean of home economics teachers.

Strong[4] estimated that the distributions of scores of home economics teachers and women-in-general overlap 32 per cent and that 2.3 per cent of the women-in-general score equal to or higher than the mean of home economics teachers. An estimate of the percentage of overlap of scores on the Strong key between distributions of home economics teachers and home economists-in-general, rather than women-in-general, would be considerably larger than 32 per cent.[5]

Summary

Items for a special vocational interest inventory in the field of home economics were selected by analyzing occupations in the field, reading vocational guidance literature, interviewing persons who had had vocational experience, analyzing data collected in a previous study, and studying general vocational interest inventories in current use.

The tentative form of the *Home Economics Interest Inventory* was mailed to 1,884 professional home economists and returned by 1,175 of them. Responses to individual items by persons in ten occupational groups were analyzed using a chi-square technique, to determine which items differentiated among occupational groups. Ninety-two per cent of the 448 items were significant at or beyond the 5 per cent level, a percentage which indicates that the technique used to select items was highly satisfactory. Items were selected for ten scoring keys and weights of 4, 3, 2, 1, and 0 were tentatively assigned to the levels of response to these items.

Other investigations in the series of studies of which the present one is a part might well include: the development of scoring keys for home economics occupations not yet studied, the development of norms for each occupational group based on the responses of another sample of professional home economists, comparison of the percentage of overlap when various methods of weighting items are used and a longitudinal study of the interests of students exploring the possibility of using their responses as the basis for developing scoring keys.

Received October 1, 1951.

[3] Percentage of overlapping was determined by expressing the mean difference in standard deviation units and referring to Table 1 in J. W. Tilton. The measurement of overlapping. *J. educ. Psychol.*, 1937, **28**, 658.

[4] Strong, E. K. Information on the mean scores of women-in-general on the home economics teacher key and overlapping between home economics teachers and women-in-general. (Private communication, June 1, 1950.)

[5] Strong, E. K. *Vocational interests of men and women.* Stanford University, California: Stanford University Press. 1943. Chapter 21.

Effect of Changing the Number of Item Responses from Five to Four in the Same Test

Ardis Swordes

State Civil Service Commission, Springfield, Illinois

During a routine analysis of our Clerical Series of 1948 an interesting situation was noticed. An analysis had been run on the Clerical Series using three groups of 100 papers each—those with highest scores, lowest passing scores, and lowest scores.

Answers on all distracters were recorded. Most of the test items had five distracters, but two sections had four, items 91–100 on Mathematics, and items 101-110 on Following Written Instructions. The results ought to have shown a zero response in the fifth space for these sections of the test, but instead a surprising number of responses appeared in the fifth space. These were found to be the result of marks that the applicants had put into the fifth space on the answer sheet.

A total count of such errors on these papers was then made as shown in Table 1. The total of all such marks was 160.

Since there were too many such marking errors to be dismissed as negligible, or the result of chance alone, an attempt was made to account for them. While the candidates were answering the first 90 questions, they perhaps developed a "set" to make the mark for the last answer in the fifth space on the answer sheet. When the group of questions with four answers was reached, the same "set" continued and the last answer, now the fourth choice, was still placed in the fifth space. In the same way the mark for the next to the last answer was placed in the fourth space occasionally, although it should have gone into the third. As might be expected, more wrong marks resulted when the correct answer was the fourth than when any of the others was correct. A tabulation of the average number of errors for each correct distracter showed the results given in Table 2.

Table 1

Total of Marks in the Fifth Space for Items with Four Responses in the Clerical Series Analysis of May 1948

Items 91-100	Clerk Typist I	Clerk Typist II	Clerk Steno. I	Clerk Steno. II	Totals
High Pass	7	4	0	4	15
Low Pass	17	6	12	12	47
Failed	18	12	9	12	51
Totals	42	22	21	28	113
Items 101-110					
High Pass	—	5	—	4	9
Low Pass	—	6	—	6	12
Failed	—	11	—	15	26
Totals		22		25	47

These results necessarily include the effect of guessing at the right answer and selecting the fourth distracter when it was wrong, and similar errors.

There were about two and one-half times as many total errors in the Low Pass group as in the High Pass group, and more than three times as many in the Failed group as in the High Pass group for Items 91–100. The High Pass group presumably was the most alert and intelligent and consequently made the fewest errors. The Failed group had the least ability to adapt to the change. This seems to penalize the low groups unnecessarily for a mechanical error. Even one point may be the difference between passing and failing.

In items 91–100 the Clerk Typist I group made the most mistakes. They have the lowest qualifications in experience and training, and should be protected from situations leading to such errors, unless choosing the correct position is considered an ability to be measured by the test.

In view of the foregoing conclusions the following steps have been taken to reduce the undesirable results of using four and five responses in the same examination. As far as possible the same number of choices is used

Table 2

Average Number of Marks in the Fifth Space for Each Correct Response to Questions 91–110

Correct Answer	Average Number of Marks in the Fifth Space
(1)	2.0
(2)	11.5
(3)	7.0
(4)	15.5

throughout the test. Special directions are given when it seems necessary to have a group of items with a different number of distracters. Questions with four answers are grouped together, instead of being alternated with items with five choices. Such precautions are especially desirable when a written test is being perpared for a position for which the qualifications required are comparatively low.

Summary

Items with four choices and items with five choices were commonly used in the same civil service examination. It was discovered that a good many applicants were placing a mark in the fifth space on the answer sheet when they answered the items with four response spaces.

A count of the total number of such marks in a written test showed a total that could hardly have occurred by chance. Technicians concluded that certain precautions should be taken to reduce the undesirable results of using a different number of distracters in the same examination. These included special instructions, reduction of the number of alternate groups, and restriction of a varying number of choices to the better qualified groups, when practicable.

Received September 28, 1951.

Some Research Findings with the Wonderlic Personnel Test

Edward N. Hay

Edward N. Hay & Associates, Inc., Philadelphia, Pa.

During 1945 each of 400 young women applicants for clerical positions in a large organization was given two forms of the Wonderlic Personnel Test, Forms D and F. This was part of a project to develop a one-minute "warm-up" test, known as Test 1.[1] In the course of this work it became apparent that there was a consistent difference in difficulty between these two forms of the Personnel Test.

In order to insure random sampling the four test sequences listed below were given in rotation. That is, every fourth applicant took the tests in order number 1, etc., as follows:

	N	*Order of Administration of Tests*
1.	100	Test 1, Personnel F, Personnel D
2.	100	Test 1, Personnel D, Personnel F
3.	100	Personnel F, Personnel D
4.	100	Personnel D, Personnel F

Table 1 shows the means and sigmas for each group of 100 cases for Form D and Form F. Since the rotating order in which the tests were given neutralizes the practice effect, the difference of 1.36 in the means is apparently

[1] Hay, E. N. A warm-up test. *Personnel Psychol.*, 1950, **3**, 221-223.

Table 1

Difference in Difficulty of Forms D and F

Order	M_D	$Sigma_D$	M_F	$Sigma_F$	$M_D - M_F$
1,FD	24.03	6.38	23.54	7.04	+ .49
1,DF	22.13	5.52	25.26	5.76	−3.13
FD	22.95	5.70	22.39	5.80	+ .56
DF	21.61	5.70	24.95	6.21	−3.34
Aver.	22.68		24.04		−1.36

Table 2

Practice Effects

Order	M_{1st}	M_{2nd}	$M_{1st} - M_{2nd}$
1,FD	23.54	24.03	− .49
1,DF	22.13	25.26	−3.13
FD	22.39	22.95	− .56
DF	21.61	24.95	−3.34
Aver.	22.42	24.30	−1.88

Table 3

Statistical Summary of Changes in Score

Change in Score	*f*
16	1
15	
14	
13	
12	2
11	1
10	5
9	6
8	8
7	19
6	22
5	20
4	34
3	48
2	55
1	37
0	48
−1	31
−2	30
−3	15
−4	9
−5	3
−6	5
−7	
−8	
−9	1
	400

Changes in Score, 1st to 2nd Test

	N	%
Increased	258	64.5%
Unchanged	48	12.0
Decreased	94	23.5
	400	100.0%

Average Score Changes

	N	*Points*
Increased	258	3.92
Decreased	94	2.40
Net Increase		1.965

a difference in difficulty of the tests themselves. In other words, Form F is 1.36 score points easier on the average than Form D. This difference is more than three times the sigma of the difference of the means and is statistically significant at the 1% level. Unpublished figures gathered by the American Bankers Association show the same approximately one point difference, in over a thousand cases.

Practice Effect

There is a significant practice effect when one form of the Personnel Test is given immediately following the other. The average increase in score of the second test over the first test is the practice effect of the first test. Any difference in difficulty of the two tests was neutralized by giving Form D first in half the cases and Form F in the other half. The results are shown in Table 2 which indicates that the average increase in score of the second test—the practice effect—is 1.88 raw score points.

To study the practice effect from another point of view, a distribution was made of the changes in score between the first and second administrations of the Personnel Test. Approximately to equalize the difference in difficulty of 1.36 points between Forms D and F, one point was added to the D scores in each case. The distribution of changes in score is shown in Table 3.

Practice Effect of Test 1

When Test 1 is not given first, the average increase in score of the second Personnel Test over the first is 1.95 raw score points. When Test 1 is given first, this practice effect is 1.81 raw score points. This difference of 0.14 score points is no doubt due to the practice effect of Test 1. In all cases the difference in difficulty of forms D and F is nuetralized by administering Form D first in one half of the cases, and Form F in the other half.

The practice effect of Test 1 is not statistically significant, perhaps because it is too short. It requires 1 minute as compared with 12 minutes for each of the Personnel Tests. Test 1 does, however, seem to serve a useful purpose in some cases by relieving nervousness and by acquainting the naive subject with the test situation.

Received March 7, 1952.
Early publication.

Speed and Accuracy of Reading Arabic and Roman Numerals*

Dallis K. Perry

University of Minnesota

It is quite obvious that Arabic numerals are read much faster and more easily in our culture than are Roman numerals. However, there has apparently been no previous attempt to determine how much faster and more accurately the Arabic numerals can be read. The present study was designed to find out how much speed and accuracy of reading are lost by the use of Roman numerals.

Procedure

Three sets of numbers, selected from a random number table,[1] were used. The first set contained numbers from one to nine, the second set contained numbers from 10 to 49, and the third set contained numbers from 50 to 99. Each set was typed on bond paper in pica type, in Arabic and in Roman numerals. The numbers were separated by spaces equivalent to two digits or letters.

The numbers were presented to subjects with the order of Arabic and Roman sets alternated in the following pattern:

	Numbers		
Subject	1–9	10–49	50–99
1	A R	R A	A R
2	R A	A R	R A
3	A R	R A	A R
etc.			

The subjects were allowed one minute for each set and were told to read aloud as fast and as accurately as possible until told to stop. The experimenter timed each reading and marked errors and the total number read on a check sheet.

The subjects were 30 senior and graduate students at the University of Minnesota. No attempt was made to obtain a random sample of any particular group, but there is no reason to believe that these subjects would have a greater familiarity with either type of numeral than any other group.

The mean number of each group of numbers read per minute and the mean number of errors per number read were calculated for both types of numerals. The significance of the differences between the means were tested, and the percentages of increase in speed and decrease in errors for Arabic over Roman numerals were calculated.

Results

Speed of Reading. Table 1 summarizes the measurements of speed of reading the two types of numerals. The differences in favor of Arabic numerals were significant at the one per cent level for all three sets of numbers. The absolute difference in number of numbers read per minute ranges from 61.4 for the single digit set to 85.0 for the 50–99 set. These differences represent 50.1, 137.5, and 349.4 per cent more numbers read per minute in Arabic than in Roman numerals.

* The writer expresses grateful acknowledgment to Profs. Donald G. Paterson and Miles A. Tinker for suggestion of the problem and for assistance in designing the experiment.

[1] Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart and Company, Inc., 1950, 378–382.

Table 1

Differences and Significances of Differences in Reading Speed* of Arabic and Roman Numerals (N = 30)

Digits	Arabic		Roman		Percentage Difference	Significance Level
	Mean	S.D.	Mean	S.D.		P
1–9	183.9	31.2	122.5	16.2	50.1	.01
10–49	115.7	19.8	40.3	9.4	137.5	.01
50–99	109.4	18.7	24.4	8.2	349.4	.01

* Numbers per minute.

Table 2

Differences and Significances of Differences in Accuracy* of Reading Arabic and Roman Numerals (N = 30)

Digits	Arabic		Roman		Percentage Difference	Significance Level
	Mean	S.D.	Mean	S.D.		P
1–9	.1	.3	.4	.9	75.0	.05
10–49	.3	.7	8.4	10.4	96.4	.01
50–99	.3	.9	10.2	9.0	97.1	.01

* Errors per 100 numbers read.

Accuracy. Table 2 summarizes the measurements of errors in reading Arabic and Roman numerals. The number of errors per 100 numbers read is significantly greater at the five per cent level for Roman than for Arabic single digit numbers. The differences in favor of Arabic numerals for the two sets of larger numbers were significant at the one per cent level. The absolute difference in errors per 100 numbers ranges from .3 for single digit numbers to 9.9 for the 50–99 set. These differences represent 75.0, 96.4, and 97.1 per cent fewer errors made with Arabic than with Roman numerals.

Discussion

That the absolute and relative differences in favor of Arabic numerals increase as the numbers get larger is not surprising; and it is probable that they would be even greater for larger numbers, which in the Roman system not only get more complex but add new symbols as well.

Roman numerals are frequently used for chapter and volume numbers, numbers of graphs and tables, dates on buildings, movies, and books, etc. The results obtained indicate that such use slows the reader and may cause errors. Such numbers are usually found in comparative isolation. Quite possibly they would cause even more hesitation and error when presented in context than in isolation as they were in this experiment.

Although the absolute and percentage differences in speed are substantial, the loss of time resulting from the slower reading of the few Roman numbers ordinarily met with will not often be large. Nevertheless, there should be more than personal preference or desire for variety to justify wasting even small amounts of the reader's time.

On the other hand, a single error may be very important. It may be assumed that, if a number is presented to a reader, it is with the intention of its being read correctly. The results show Roman numerals to be far less satisfactory than Arabic in this respect.

Summary

1. An experiment was carried out to determine how much faster and more accurately Arabic numerals are read than Roman numerals.

2. Thirty college students were asked to read as fast and as accurately as possible sets of numbers from one to nine, 10 to 49, and 50 to 99 in both Arabic and Roman numerals, and measurements of speed and errors were taken.

3. Percentages of increase in speed of reading for Arabic over Roman numerals were 50.1 per cent for numbers from one to nine, 137.5 per cent for numbers from 10 to 49, and 349.4 per cent for numbers from 50 to 99. All differences were significant at the one per cent level.

4. Percentages of decrease in errors for Arabic from Roman numerals were 75 per cent for numbers from one to nine, 96.4 per cent for numbers from 10 to 49, and 97.1 per cent for numbers from 50 to 99. The first difference was significant at the five per cent level and the last two at the one per cent level.

5. Use of Arabic rather than Roman numerals would seem to be desirable for most purposes because the former are read faster and more accurately.

Received November 5, 1951.

The Dimensional Analysis of Motion: IV. Transfer Effects and Direction of Movement[1]

Patricia von Trebra and Karl U. Smith

University of Wisconsin

The dimensional analysis of human motions is based on three main principles which represent a departure from the current concepts of investigation of motor coordination, especially those identified with the study of psychomotor skills and of time and motion analysis. These three principles are: (1) separate measurement of the unrelated component movements of a motion pattern; (2) development and use of pre-planned performance situations that permit explicit quantitative control of the reactive dimensions of motion; and (3) investigation of the determining variables affecting motion in relation to known and specifiable quantitative reactive dimensions that identify the movement pattern. New techniques, to be described below, permit implementation of these principles for the purposes of the present study.

In the investigation to be reported here, a dimensional study of transfer of training has been conducted. The degree of transfer from one pattern of motion to others has been investigated relative to controlled change in the direction of movement in the motion pattern. The transfer effects measured cover the manipulative and travel components of movement in the motion investigated. At the same time, information has been obtained concerning the nature of learning different directions of manual motions.

Methods

1. *Apparatus.* The main objectives of the present study have been to measure separately the unrelated travel and manipulative components of a complex motion pattern when the subject is trained in a particular direction of movement and is thereafter required to perform the same pattern of movement in different directions of motion.

In order to segregate the travel and manipulative components of the motion pattern for measurement, the Universal Motion Analyzer (2), shown diagrammatically in Figure 1, has been used. This device consists of a special electronic relay circuit that will measure automatically the duration of different components of motion. The relay is connected to the subject, to the work situation, and to two precision time clocks. The device operates in such a way that the subject's contact with objects or controls to be manipulated acts to key the relay and thus activate the clocks. A dual operation is involved in the relay circuits. The duration of time in which the subject is in contact with the work object or control is registered on the manipulation time clock. The duration of time required to travel between two work objects or controls is registered on the travel time clock. Time is measured in .01 seconds. The clocks used are calibrated to a precision of .005 seconds.

Experimental control of the reactive dimensions of movement is accomplished by means of pre-planned performance situations that permit quantitative variation of the reactive space dimensions of the motion pattern. The performance situation used here is a large control panel equipped with turn-switches, as shown in Figure 1. These switches, and the control panel on which they are arranged, are designed so that the direction of motion, the pattern of manipulation, the plane of movement, the distance of travel, the complexity of movement, and other dimensions defining the motion pattern may be controlled as desired.

In order to vary the direction of motion, the four patterns of movement shown in Figure 2A

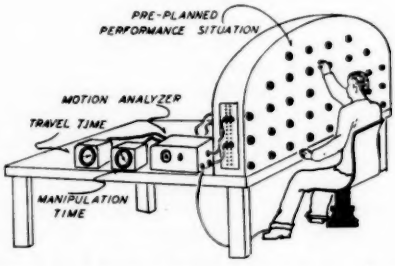

FIG. 1. Diagram of the pre-planned performance situation with the Universal Motion Analyzer.

[1] This research has been supported by funds granted by the Graduate School Research Committee, University of Wisconsin.

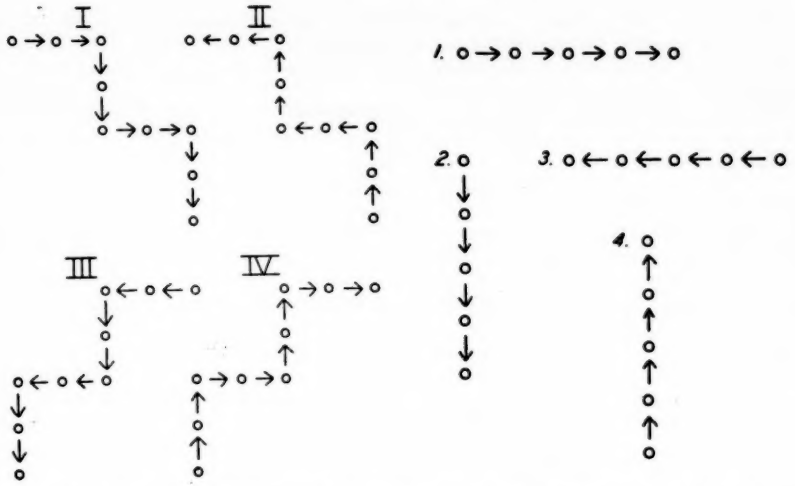

FIG. 2. The patterns of motion used in the transfer tests (2A) and in the pretraining tests (2B).

were used. The small circles represent the turn-switches operated, and the arrows indicate the direction of motion between switches. Pattern 1 is a right-down movement. The other three patterns are a left-up motion (Pattern II), a left-down motion (Pattern III) and a right-up motion (Pattern IV). Each pattern of movement is identical except for the direction of travel within the total motion. There are nine manipulations in each pattern.

As already described, the subjects of this experiment were required to practice one pattern of motion and were thereafter tested in their performance on other directions of movement. To secure an appropriate experimental design of the observations, the subjects used were divided into four matched groups on the basis of their level of performance in a pretest with four different patterns of motion. The four motion patterns used in this pretest are indicated in Figure 2B. During the pretest, and in the main observations, the pattern of motion to be performed was indicated by a lined guide marked out on the instrument panel. All the main performances were carried out with the right hand.

2. *Experimental Design.* The subjects consisted of 48 right-handed college students, both male and female. In the procedural design, number of training days was held constant for all subjects. After training on a particular pattern of movement, the subjects were then tested on the other three patterns of motion, and the presentation of these transfer patterns was randomized by subjects. The effects of individual differences were kept at a minimum by the use of four matched groups, which were formed in terms of performance on the pretest motion patterns. The experimental design is summarized in Table 1. Details of this design may be indicated as follows:

a. In order to get four matched groups, all subjects were pretrained over a period of 4 trials on each of the four pretraining patterns described in Figure 2B. These patterns were selected because they approximated the directional characteristics involved in the experimental work situation without the directional complications of the training patterns. Each subject performed the entire 16-trial sequence in random order, first with his right hand, then with his left hand.

b. To provide training over an extended period of time, each group was trained in one of the four training patterns. Each subject received 4 trials per day for 8 successive days on his particular group pattern.

c. To obtain a measure of transfer effect, each subject performed one trial on each of the four movement patterns, including the pattern on which he had been trained, on the day following completion of training. There were four patterns of transfer order which were randomized by subjects within each group so that each pattern of 4 trials was performed by 3 of the 12 subjects in each group.

Table 1

Experimental Design

Pretraining Trials[1]	Training Conditions[2]	Sub. No.	Training Days[3]	Transfer Conditions[4]
1 4 2 3	I	1	1 2 3 4 5 6 7 8	A I II III IV
4 3 1 2		.		B IV III I II
3 2 4 1		.		C III II IV I
2 1 3 4		12	– – – – – – – –	D II I III IV
	II	13		
		.		
		24		
	III	25		
		.		
		36		
	IV	37		
		.		
		48	– – – – – – – –	

[1] This sequence of trials, performed with both right and left hand, involves performance on four patterns, randomized as shown. The same sequence was used for all subjects.

[2] Subjects were divided into four matched groups on the basis of the mean right-hand performance time in the pretraining.

[3] Four trials were given each day with the right hand, making 32 practice trials in all.

[4] The four different orders were randomized by subjects. Each subject performs in one series of trials the conditions indicated.

Results

The results of this experiment may be divided into three general parts for presentation. Learning effects found in the experiment will be described first, then the data related to the transfer observations. Finally, some special observations related to transfer effects will be noted.

1. *Learning Effects.* During the eight days' practice by the four different groups on each of their respective patterns of motion, learning differences were found for different directional patterns and for the two component aspects of motion measured. Although these data have not been subjected to detailed statistical analysis, it may be noted that the manipulation component of the motion pattern showed a progressive learning effect throughout the eight days' practice for all four directions of motion. In contrast, the travel movements showed inconsistent changes related to learning. There is no significant learning effect for travel movements on Pattern IV by Group IV. This is the right-up pattern of movement. The over-all change in the travel component of the movements due to practice is very small for those patterns giving a learning effect at all. This change amounts roughly to nine per cent decrease in time over the eight days. The mean decrease in duration of manipulation due to practice is around fifteen per cent. The learning effect on the travel components of the movements takes place over the first three or four days, if it occurs at all, whereas the effect of practice on manipulation is uniform and continuous over all eight days.

The relative efficiency of performance of the different groups varies during training, although the differences are not large. Manipulative performance by Groups I and III on their respective patterns is inferior to that of Groups

II and IV on their learning patterns. The most efficient of all groups in manipulation during training is Group II on Pattern II, the left-up motion.

In terms of performance of the travel component of movement during learning, Groups I and III on their respective patterns are the most efficient, and Groups II and IV are the least efficient. Thus, the group and patterns showing consistently the fastest manipulative motions during training display slower travel motions.

Transfer Effects. Four separate analyses of variance were computed for the data concerning performance of the two hands and the manipulation and travel data for each hand.[2] These analyses brought out two main facts. Differences due to training groups were shown in the transfer tests to have no statistical significance at all. This statistical test may be said to indicate, among other things, that the initial matching of the training groups was successful. However, differences in performance due to the transfer patterns were significant in all four analyses. That is to say, both manipulation and travel time data for each hand show significant variation as a result of the presentation of different patterns of movement during transfer. These differences due to the effects of transfer were significant at the .01 point in all but the analysis of variance in left-hand travel time. In this case, transfer patterns produced differences which were significant at the .05 point.

The effect of order of performing the different patterns of movement in the transfer test is also significant in this experiment. As mentioned before, this order was randomized for subjects within each group.

The actual phenomena of transfer of learned response observed in this experiment may be described by considering the degree to which performance on a given pattern of movement will be changed, either positively or negatively, by training upon other patterns of motion. To express this change quantitatively, the difference between initial performance on a given pattern and the performance on that pattern during the transfer test has been computed. To obtain a "transfer index," this difference is then divided by the value representing the initial level of performance and the ratio multiplied by 100.

Figure 3 summarizes the mean transfer index for the four different transfer patterns. These values are described in a line graph even though they do not represent any sort of continuous function. This graph is obtained by averaging the three transfer indices of three different patterns of movement on which different subject groups were trained with reference to a particular pattern to which these three training groups transferred. The graph may be interpreted to indicate which of the four patterns of movement gave the greatest amount of positive or negative transfer for manipulation and travel components of the motion pattern

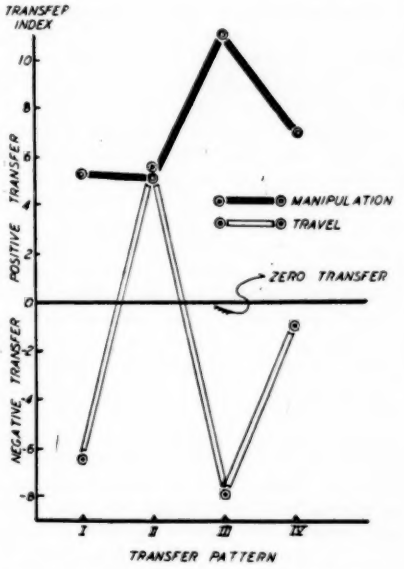

FIG. 3. The variation in the level of transfer from three other directional patterns of motion to a given directional pattern for manipulation (solid line) and travel movements (open line). Level of transfer is indicated in terms of an index.

[2] The critical figures of this experiment and the summaries of the analyses of variance are on file with the American Documentation Institute. Order Document 3420 from American Documentation Institute, 1719 N Street N.W., Washington 6, D. C., remitting $1.00 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $1.00 for photocopies (6 × 8 inches) readable without optical aid.

when training has been carried out in three other different directional patterns.

Figure 3 shows that the manipulation component of the movement pattern displays positive transfer effects on all four transfer patterns used. This means that practice on any of the training patterns will produce an improvement in performance in manipulation on any one of the transfer patterns. In contrast to this, training on given patterns of movement produces generally a decrement in performance or a negative transfer effect for the travel component of the motion pattern. Transfer to Pattern II is the only pattern of the four used which displays a positive transfer effect for the travel component of motion.

For manipulation, the pattern showing the largest transfer effect in terms of these mean values is Pattern III. In Pattern III, the subject moves his arm down and to the left.

These data on transfer effects may also be described by defining the training patterns which produce the greatest amounts of positive or negative transfer to other patterns of movement. These relations are very complex and difficult to describe in the limited space available. When the data are viewed in these terms, however, it is seen again that the transfer effects on the two different components of movement are quite distinct. The transfer effects for manipulation are generally positive for different training patterns, whereas these effects for travel are generally negative. Pattern II, in which the direction of movement is up and to the left, generally produces a higher positive transfer effect in manipulation to the other three patterns. The travel component of motion shows the highest negative transfer effects in the case of transfer from Pattern IV to Patterns I and III. In Pattern IV the subject moves up and to the right. The most consistent negative-transfer effects are found in transfer from Pattern II to all other patterns.

The discussion above has centered around the relative level of performance of the right hand in the transfer tests as compared to initial performance with this hand. The design of this experiment also permitted some estimation of the relative level of proficiency of the left-hand performance in manipulation and travel on the different transfer patterns.

Bilateral transfer from right hand to left hand, when training covers all different patterns, produces the most rapid performance in manipulation on Patterns II and III. Similar transfer effects in the travel component of movement produces the most rapid performance on Pattern III.

The training patterns of the right hand associated with the most rapid performance in transfer of manipulation to the left hand are Patterns II and III. In travel movements, training on Pattern III produces the most efficient performance on all transfer patterns with the left hand.

In addition to the above observations, some data were also obtained on the persistence of transfer effects. These observations were made by repeating the transfer tests on Group I at an interval of one week after the initial transfer tests. The relative and absolute levels of performance on different transfer patterns, and on the training patterns for this group, showed no marked changes in manipulation or travel after this one-week interval. In other words, the effects of learning a particular pattern of motion upon performance in other directions of movement apparently persist without significant alteration over an interval of at least one week.

Summary

A new systematic point of view, along with pertinent specific principles, has been described for the applied and theoretical study of motion. Special new techniques, implementing these principles, have been used in this investigation of transfer of learned movements from one direction of motion to others.

Learning data indicate that, for different directional patterns of motion, the travel component on the pattern shows limited and inconsistent changes with practice, whereas the manipulative component of the pattern displays uniform progressive change throughout the eight days of training. These data are consistent with earlier findings in showing that learning effects for different component movements in a motion pattern are quite distinct (**1, 3**).

Transfer of response from one directional pattern of motion to others presents certain striking features when examined in terms of the component movements of the motion sequence.

Manipulative movements show generally a positive transfer effect. Travel movements display generally negative transfer effects. Significant effects exist between the directional change and both components of motion even though the direction of travel is the only dimension of motion which is altered experimentally.

Generally speaking, some directions of motion show a greater transfer effect than others when training is carried out on the other different directions of motion. Similarly, the degree of transfer effect from a given training pattern varies with the direction of response. All of these effects are specific to the component of movement under consideration.

Bilateral transfer effects occur. Generally speaking, there are two patterns of motion, among the four studied, which are associated with the high levels of efficiency by the left hand when transfer occurs bilaterally.

Transfer effects, as described, persist, relatively unaltered, over an interval of one week's time.

The phenomenon of transfer of training is a fundamental problem in the applied psychology of training, training equipment design, instrumentation, and planning of work. Heretofore, general notions about transfer effects in these fields have been guided by general theoretical points of view identified by the terms, identical elements, transposition and response induction. The results presented here not only define the limitations in application of such generalized ideas but also prescribe the systematic empirical orientation necessary to specify and predict quantitatively the detailed phenomena of motor coordination where the human subject must change a learned mode of response. Technical developments in the quantitative measurement of component movements in the motion pattern and in the design of pre-planned quantitatively controlled task situations are fundamental to this empirical approach to the study of the psychophysical, psychophysiological, and dynamic phenomena of motor coordination.

Received October 9, 1951.

References

1. Rubin, G., von Trebra, Patricia J., and Smith, K. U. Dimensional analysis of motion. III: Complexity of movement pattern. *J. appl. Psychol.*, 1952, **36**, 272–276.
2. Smith, K. U., and Wehrkamp, R. A. Universal motion analyzer applied to psychomotor performance. *Science*, 1951, **113**, 242–244.
3. Wehrkamp, R. A., and Smith, K. U. Dimensional analysis of motion. II: Travel distance effects. *J. appl. Psychol.*, 1936, **36**, 201–206.

Book Reviews

Rose, Arnold M. *Union solidarity*. Minneapolis: Univ. Minnesota Press, 1952. Pp. 209. $3.00.

Union Solidarity is an unusual work in many respects. It presents one of the few research efforts concerning the internal workings of unions. Moreover, the research was initiated by a union, a type of organization which has been very reluctant to bare its intimacies to public survey. But perhaps most surprising of all, it reports findings using actual (not disguised for purposes of anonymity) names and places.

The title of Dr. Rose's book may be to some a bit misleading. It is not a work of an armchair academician, theorizing on how the "union movement" should operate to be in accord with preconceived postulates. *Union Solidarity* is, rather, a descriptive report based upon hard, cold data provided by those who know the actual significance of unionism best—the union members. Dr. Rose's book does not pretend to present the reader with a neat package of information with which to understand the intricacies of the American Labor Movement's internal workings.

Union Solidarity will provide the reader, however, with intimate details about how one local union's membership views its union. It will provide him with the possible significance of such opinion within a rather broad and somewhat vague conceptual framework of "union solidarity."

Briefly, Dr. Rose's book is based upon material gathered by means of a structured response category questionnaire interview. Although he used relatively unskilled interviewers (students subjected to a brief training session), the possible ill-effects should have been alleviated to some degree by the design which, for the most part, gave the interviewees predetermined response category alternatives as the answers to the questions. In a few questions, however, the interviewer and interviewee were given more freedom. To that extent, the reliability of the data is open to question.

Although the results are presented in terms of comparative percentages with considerable cross comparisons between categories of response, particularly with reference to the factual information questions, e.g., age, length of union membership, number of meetings attended, etc., it is difficult to see how his data could be subjected to rigorous statistical analysis. Purely in the sense of an aside, the statistically oriented reader may be somewhat puzzled by the use of the term "statistically significant" without any indication of the level of significance or how such significances were derived.

Basically, Dr. Rose has presented a descriptive work. There was little attempt to relate the interview findings to membership behavior in the union situation. His only attempt was to measure certain select categories of interview findings against other categories which, on an a priori basis, were used as criteria of union solidarity. At best, the criteria appear to be relatively vague and founded primarily on theory, not fact. In general, the study, in and of itself, appears to have little predictive potential.

At this point it must be stressed, however, that Dr. Rose's inferences, drawn from the results, are notably cautious. He is careful to state the limitations of the study, and except for the relatively definitely stated criteria of union solidarity, does not force the results to fit into any preconceived theory.

To a great extent, the reader is given the results, given several tentative explanations, and thereafter allowed to forge ahead for himself if he is desirous of more definite interpretations. Dr. Rose does proffer several generalizations, both practical and theoretical, in his last chapter. However, in keeping with the major part of his writing, he makes them more in terms of suggestions than hard and fast conclusions.

Even though *Union Solidarity* is not, and does not pretend to be, more than a descriptive evaluation of one local union, it will provide for those interested, even tangentially, in unionism, much interesting and informative data. For the union leader, it should give

insight into the potential usefulness of social science techniques; for the teacher and student of labor and industrial relations, an opportunity to get a glimpse into that elusive and hitherto relatively unexplored area, the union; and for the research worker in the field of unionism, insight into the advantages and disadvantages inherent in conducting research utilizing a comparable technique, as well as comparative data against which to relate his findings.

Hjalmar Rosen

Institute of Labor and Industrial Relations, University of Illinois

Di Michael, Salvatore G. (Ed.) *Vocational rehabilitation of the mentally retarded.* Washington: U. S. Government Printing Office, Federal Security Agency, Office of Vocational Rehabilitation, 1950. Pp. viii + 184. 45¢.

This concise yet comprehensive treatment of the vocational rehabilitation of the mentally retarded is a volume of value to those who are concerned with the adjustment of others. It is written primarily to meet the needs of the professional staff of the State-Federal program of Vocational Rehabilitation and is one of a series of staff development aids pointed toward specific rehabilitation problems.

The first six chapters of the bulletin cover some basic considerations in rehabilitation and were written by individuals who have worked intensively with the mentally retarded. The last three chapters, or second part of the bulletin, contain descriptions of three recent programs of rehabilitation of this group.

Jervis' chapter on the medical aspects of mental deficiency is one of the most succinct the reviewer has seen. It will serve as a ready reference for an operational understanding of the etiology and symptoms of mental retardation. Hegge's point of reference for the determination of the feasibility of training a client is a healthy one. This determination is made on the basis of the client's ability to benefit from the service. The need for careful avoidance of rule-of-thumb methods of selection and the importance of use of individual appraisal methods is well emphasized. The implication this carries for the necessity of having well trained and carefully chosen rehabilitation counselors might have been stressed by the writer.

In a chapter on the counseling of the mentally retarded, Yepsen emphasizes the fact that the mentally retarded are not qualitatively different from the "normal" group, but are different only in degree. He recommends the usual counseling techniques but emphasizes the need for working more slowly and for carefully appraising the counselee's understanding of each step. If there is any research particularly of the kind in which recordings have been made of counseling interviews with the mentally retarded, reference to it would be valuable. If there is none, certainly it is needed for appraisal of counseling techniques with this group.

The use of job analysis is recommended in the chapter on employment. The availability of such assistance from the USES might well have been mentioned.

It is gratifying to read the reports on the work that has actually been done in the rehabilitation of the mentally retarded. The emphasis on the job levels at which these people can succeed, the need for careful follow-up, the aid to the parents, the training in personal habits, and the assistance needed to establish good relationships with other employees are all necessary parts of an integrated program.

Almost anyone who works with people at some time encounters the mentally retarded. The suggestions found in this bulletin can be of invaluable assistance in such instances.

Vivian H. Hewer

Student Counseling Bureau, University of Minnesota

Feingold, S. Norman. *Scholarships, fellowships, and loans.* (Vol. II.) Boston: Bellman Publishing Company, Inc., 1951. Pp. 312. $5.00.

Additional information about student aids administered by some 245 agencies can be found in Vol. II of *Scholarships, Fellowships, and Loans.* Written for the student who may not have the financial resources to continue his education—as well as for his counselor—this directory lists opportunities not usually found

in school or college catalogues nor included in Vol. I.

The directory is divided into three parts: first, there is a short introductory and editorial portion; followed by brief summaries of various career aids, some 240 pages in length; and concluding with a bibliography and three indexes.

In Part I, the author summarizes the need for student aids; at the same time pointing to the failure of guidance counselors to utilize available resources because they are not generally aware of their existence. Local and regional sources, particularly, are neglected and suggestions and examples are given so that the professional worker may uncover them.

A sample application blank, hints to candidates for student aid and advice on career planning is included in this first part. Directed specifically to the student, the material on career planning is short and somewhat superficial. A check list of "things to remember" when applying for financial assistance is pertinent, well done and should help the student organize his scholarship hunt.

Page length summaries of available aids constitute the subject matter of Part II. Alphabetically arranged by sponsoring agency, information has been noted on the name and address of the resource, sex and qualifications of the applicant, amount of funds available, special fields of interest and where to obtain further information and application blanks. Whether the aid is available for undergraduate study, graduate work or research is noted in each summary. Most of the aids listed are sponsored by private organizations, rather than schools, and so are not usually found in school or college catalogues.

Part III contains a comprehensive and lengthy bibliography which should be of help to those interested in setting up a student file. Contained also are three indexes designed to make the material in the book accessible to the reader. The first index covers the introductory material. The other two indexes cover the student aid material; one is an alphabetically arranged subject index, and the other is arranged on the basis of vocational goal and field of interest.

Vol. II is intended as a companion for Vol. I and not as a substitute. Some improvements have been made in the second volume, notably that of including more general education aids on a regional or local level. The indexing of this important material could have been improved for easier accessibility.

The reviewer was impressed with the number of aids uncovered by Dr. Feingold, a native of Boston, in his own New England region. It would be a Herculean task to gather such complete information for every geographic area and the author makes no claim to having done so. He suggests—and counselors can profit from his example—the gathering of such information on a local or regional basis.

Scholarships, Fellowships, and Loans, Vol. II, together with Vol. I, offers the counselor a comprehensive, though not all-inclusive, survey of financial aids available in this country. Those working with young people can make good use of such a centralized listing.

Solomon Shapiro
Jewish Vocational Service, St. Paul, Minnesota

Berdie, R. F. (Ed.) *Concepts and programs of counseling.* Minnesota Studies in Student Personnel Work, No. 1. Minneapolis: University of Minnesota Press, 1951. Pp. 81. $1.75.

This book contains five papers originally delivered at the conference for administrators of counseling programs, held at the University of Minnesota in November, 1950. Despite their having originally been written for such a group, they should, because of their wide implications, thoughtful approach to practical problems, and largely non-technical language, be of considerable interest to the many people whose work brings them into direct or indirect contact with professional counseling.

Included are papers by O. H. Mowrer on anxiety theory and the distinction between counseling and therapy, F. M. Fletcher concerning some problems of counseling personnel, W. M. Gilbert outlining some suggestions for intra-university organizational relationships, J. L. Holmes dealing with recent developments in counseling, and P. L. Dressel considering some issues involved in counseling evaluation. Each paper was intended, originally, to form the basis for group discussion; together, they should provoke much productive thought.

Despite the diversity of topics, the underlying unity will be readily apparent, and the book is well balanced by the consideration, from different points of view, of many important, practical problems.

It is impossible to consider in detail the many highly stimulating points brought out. Mowrer's discussion, in which he proposes a distinction between counseling and therapy based on learning and personality theory rather than bias, is most interesting although it seems to imply a dichotomy, normal versus neurotic, which is hard to accept. Fletcher's incisive comments on a wide range of problems associated with counseling personnel are excellent. In large part, he presents such problems without answers, for there are none, but his relating them and opening them up should bring about more effort in this direction. Gilbert's paper is most interesting in its emphasis on certain practical—and fundamental—issues involved in intra-university relationships. His discussion of the confidentiality of counseling information is particularly worth while, even though some may consider his position extreme. Holmes' summary of recent developments highlights the relatively recent emergence of counseling as a recognizable profession. There seems, however, to be too heavy an emphasis on projective techniques to the exculsion of other equally interesting developments. One of the most stimulating papers is that by Dressel on evaluation. He not only sees the need for such work but also the many difficulties besetting those intrepid enough to undertake it. His suggestions should be welcomed by all whose professional consciences are bothering them about the state of affairs in evaluation.

If there is any over-all criticism, it would have to be that too few topics are dealt with at real depth. But perhaps the very extensiveness of coverage which, perforce, necessitated some superficiality is an *advantage*. For these papers are excellent stimuli to thought and discussion; the fact that they open up so *many* problems makes them, in this sense, particularly valuable. Dr. Berdie is to be congratulated for having planned this conference about such worth-while topics; the various writers should be commended—and read—for thoughtful, practical, and most interesting pieces of work.

John W. Gustad

University of Maryland

Keys, A., Brozek, J., Henschel, A., Mickelsen, O., and Taylor, H. L. *The biology of human starvation*. Minneapolis: University of Minnesota Press, 1950. Pp. xxxii + 1385, 2 Vol. $24.00.

These two volumes present in systematic detail not only the results of the writers' extensive and intensive study of controlled semi-starvation but also a comprehensive survey of the world's literature through 1949 dealing with the general subject.

The "Minnesota Starvation Experiment" consisted of a careful longitudinal study of the physical and psychological characteristics of 32 conscientious objectors who were subjected to semi-starvation. Thirty-six subjects were selected from 100 volunteers. (Special attention is given to the four who did not complete the experiment.) The subjects were studied throughout a 12-week control period, 24 weeks of semi-starvation (average loss of 24% of initial body weight), 12 weeks of restricted rehabilitation, 8 weeks of unrestricted rehabilitation and in some cases follow-up periods of 8–12 months.

The book is divided into six sections: background, morphology, biochemistry, physiology, psychology and special problems. Each section presents in detail the information in the literature and the new experimental data, critically reviewing theory and fact. Appendixes on methods, detailed data (119 pages), wartime diets and rations, and notable famines plus references (87 pages) and a general index complete the book.

The importance of the work to those concerned with human nutrition can hardly be overemphasized but it should be noted that it is a stimulating source of facts for the psychologists as well. The need for understanding the psychological side of starvation is recurrently stressed, e.g. "Two impressions dominate the picture; the immense importance of the psychological aspect of inanition and the comparative simplicity of the nutritional and biochemical problem," page xiii, preface.

Of particular interest to applied, clinical

and differential psychologists are the sections dealing with changes in body types, glandular activity, increased auditory acuity, drastically decreased work capacity and reduced sexual activity. Data from intelligence tests, the MMPI, GAMIN, STDCR, interviews, diaries, self ratings and autobiographies give a dramatic picture of "semi-starvation neurosis." This personality disturbance, one of the most important results on the psychological side, was characterized by apathy, irritability, introversion, depression and hysteria. It did not involve any loss of intellective capacity as measured by the CAVD and other tests. The remarkable persistence of this neurotic pattern far into rehabilitation reveals sharply some of the problems to be faced by the rehabilitators.

The scope, coherence and precision of the work are impressive. These volumes constitute an important reference source for much data, new and old, and offer hypotheses and suggestions which should challenge psychologists.

James J. Jenkins
University of Minnesota

Jalota, S. *Scientific personnel selection procedure: A study.* Hind Art Press, Godowlia, India.

This little monograph describes the personnel selection procedures used by the Civil Selection Board and the Tata Iron and Steel Co. of India. Candidates for positions are evaluated by psychologists' judgments of fitness based principally upon intelligence tests and personality measures of various sorts. These judgments are "validated" against the ratings of a final board which considered the psychologists' reports together with other information concerning the candidates, such as previous experience.

The American psychologist will find little to interest him in this monograph either by way of methodology or results. He will be disappointed to find such non-standardized evaluative procedures and primitive validation techniques being utilized. The extensive tables giving distribution statistics of and intercorrelations among ratings serve little purpose. Yet the style of writing has such charm and the author is so facile at turning a phrase that the reviewer could not help being intrigued. Certain errors are described as negligible because they do not cost the psychologist any anxiety or sleepless nights. Reference is made to the Rorschach technique as a "closed preserve." Personality inventories are said to be of little value because "To ask a candidate to condemn himself for a specific job is too utopian a hope for its fulfilment in modern hard, materialistic times." One can only hope that a writer with such pleasing expression will in the future have a message of some value to present.

Edwin E. Ghiselli
University of California

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

Work measurement, new principles and procedures. Adam Abruzzi. New York: Columbia University Press, 1952. Pp. 290. $6.00.

Dynamic psychiatry. Franz Alexander and Helen Ross, Editors. Chicago: University of Chicago Press, 1952. Pp. 578. $10.00.

Child Rorschach responses. Louise Bates Ames *et al.* New York: Paul B. Hoeber, Inc., 1952. Pp. 310. $7.50.

Management controls in industrial research organizations. Robert N. Anthony. Boston: Harvard Business School, 1952. Pp. 538. $6.75.

The Arthur adaptation of the Leiter International Performance Scale. Grace Arthur. Washington, D. C.: The Psychological Service Center Press, 1952. Pp. 73. $3.00.

Social psychology. Solomon Asch. New York: Prentice-Hall, Inc., 1952. Pp. 736.

Equality by statute. Monroe Berger. New York: Columbia University Press, 1952. Pp. 238. $3.25.

Practical psychology. Revised edition. F. K. Berrien. New York: The Macmillan Co., 1952. Pp. 640. $5.00.

Disorganization, personal and social. Herbert A. Bloch. New York: Alfred A. Knopf, 1952. Pp. 608. $5.00.

Marketing research. Ernest S. Bradford. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 379. $5.00.

The people elect a president. Angus Campbell and Robert L. Kahn. Ann Arbor: University of Michigan Press, 1952. $1.50.

The selection of military manpower. Leonard Carmichael and Leonard C. Mead, Editors. Washington, D. C.: National Research Council, 1952. Pp. 270. $2.50.

Cruelty to children. Eustace Chesser. New York: Philosophical Library, Inc., 1952. Pp. 159. $3.75.

How to supervise people. Alfred M. Cooper. New York: McGraw-Hill Book Co., 1952. Pp. 254. $3.75.

'Twixt the cup and the lip. Margaret Cussler and Mary L. de Give. New York: Twayne Publishers, 1952. Pp. 262. $3.95.

Planning and developing the company organization structure. Ernest Dale. New York: American Management Association, 1952. Pp. 232. $4.50.

Elementary school guidance. Ervin Winfred Detjen and Mary Ford Detjen. New York: McGraw-Hill Book Co., 1952. Pp. 266. $3.75.

Interracial housing: a psychological evaluation of a social experiment. Morton Deutsch and Mary Evans Collins. Minneapolis: University of Minnesota Press, 1952. $3.00.

Social psychology. Leonard W. Doob. New York: Henry Holt and Co., 1952. Pp. 583. $5.00.

A dictionary of psychology. James Drever. Baltimore: Penguin Books, Inc., 1952. Pp. 315. $.85.

Measurements of human behavior. Revised edition. Edward B. Greene. New York: The Odyssey Press, 1952. Pp. 790. $4.75.

Fundamentals of social psychology. Eugene L. Hartley and Ruth E. Hartley. New York: Alfred A. Knopf, 1952. Pp. 740. $5.50.

Personnel administration and labor relations. Herbert G. Heneman and John G. Turnbull. New York: Prentice-Hall, Inc., 1952. Pp. 444. $3.95.

The origin of life and the evolution of living things. Olan R. Hyndman. New York: Philosophical Library, 1952. Pp. 648. $8.75.

The adolescent and his world. Irene M. Josselyn. New York: Family Service Association of America, 1952. Pp. 124. $1.75.

Color in business, science and industry. Deane B. Judd. New York: John Wiley and Sons, Inc., 1952. Pp. 401. $6.50.

Readings in industrial and business psychology. Harry W. Karn and B. von Haller Gilmer. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 476. $4.50.

Personality and conflict in Jamaica. Madeline Kerr. Liverpool: University Press of Liverpool, 1952. Pp. 221. 15s.

Psychoanalytic explorations in art. Ernst Kris. New York: International Universities Press, 1952. Pp. 358. $7.50.

Adolescence. Marguerite Malm and Olis G. Jamison. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 512. $5.00.

AFL attitudes toward production, 1900–1932. Jean Trepp McKelvey. Ithaca: New York State School of Industrial and Labor Relations, Cornell University, 1952. Pp. 160. $1.00.

Your complexes and you. Vitali Negri. Los Angeles: American Book Institute, 1952. Pp. 299. $3.50

Dictionary of mind, matter and morals. Bertrand Russell. New York: Philosophical Library, 1952. Pp. 290. $5.00.

Understanding ourselves. Helen Shacter. Bloomington: McKnight and McKnight Publishing Co., 1952. Pp. 124. $.70.

Psychology. Ross Stagner and T. F. Karwoski. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 582. $5.00.

Measuring your public relations. Herman D. Stein. New York: National Publicity Council for Health and Welfare Services, 1952. Pp. 48. $1.25.

Employment of the older worker. Clark Tibbitts, Arthur J. Noetzel, Jr., and Charles C. Gibbons. Kalamazoo: W. E. Upjohn Institute for Community Research, 1952. Pp. 24. Gratis.

Applied statistics, a journal of the Royal Statistical Society. Leonard H. C. Tippett, Editor. London: Oliver and Boyd Ltd., 1952. $4.00 per year.

Handwriting. Frank Victor. Springfield: Charles C Thomas, Publisher, 1952. Pp. 149. $3.75.

The man on the assembly line. Charles R. Walker and Robert H. Guest. Cambridge: Harvard University Press, 1952. Pp. 180. $3.25.

The clinical method in psychology. Robert I. Watson. New York: Harper and Brothers, 1951. Pp. 779. $5.00.

Elements of statistical method. Third edition. Albert E. Waugh. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 531. $5.50.

The dream—mirror of conscience. Werner Wolff. New York: Grune and Stratton, Inc., 1952. Pp. 348. $8.50.

Joint consultation in British industry. National Institute of Industrial Psychology. London: Staples Press Limited, 1952. Pp. 276. 21s.

Morale—and the prevention and control of panic. New York: New York Academy of Medicine and Josiah Macy, Jr. Foundation, 1952. Pp. 75.

The Psychological Corporation 30th Annual Report. New York: The Psychological Corporation, 1951. Pp. 16. Gratis.

WILEY

What happens when psychological principles are actually put to work with industry as the laboratory?

Read . . .

PRINCIPLES of HUMAN RELATIONS

By NORMAN R.F. MAIER

Professor of Psychology
University of Michigan

Drawing from social psychology, group dynamics, clinical psychology, and the other fields in which human relations have been studied, Dr. Maier applied these principles to existing industrial organizations. The result is the establishment of truly effective methods for (1) overcoming communication barriers, (2) preventing misunderstanding, and (3) developing the constructive side of man's nature. The examples and applications are pointed toward industry, but the principles apply to all situations where leaders must deal with groups or individuals.

This comment from

Professor D.J. Moffie
Head, Department of Psychology
University of North Carolina

". . . very timely, scientific in scope, and what is now needed in industry. Considerable emphasis in the past years has been placed on role taking, human relations, and the conference method. This book, I believe, will give industry some concrete basis for this field of study."

And from

Dr. B. von Haller Gilmer
Department of Psychology and Education
Carnegie Institute of Technology

"Dr. Maier has combined sound, fundamental principles of psychology with practical suggestions of implementation . . . it is encouraging to have a sound scholar write such a practical book . . . congratulations to Wiley and Dr. Maier for giving the field of industrial psychology such an important book."

See group decision, non-directive counseling, "role-playing," the conference method, and many other tested and proven techniques in action. Send today for an on-approval copy of Maier's *Principles of Human Relations.*

1952 **474 pages** **$6.00**

JOHN WILEY & SONS, Inc., 440 Fourth Ave., New York 16, N.Y.

Distinctive McGRAW-HILL *Books*

PRACTICAL PSYCHOLOGY. New 2nd Edition

By KARL S. BERNHARDT, University of Toronto. Ready in January

Designed for students who are doing work at a post secondary school level, this text provides a sound treatment of psychology using a practical approach. Simple in treatment with down-to-earth applications, *Practical Psychology* covers most of the usual topics in introductory psychology, but adds chapters on child training, mental hygiene and personal adjustment. A chapter on aptitude measurement and another on personal happiness and efficiency have been added.

THE EMPLOYMENT INTERVIEW IN INDUSTRY

By NEWELL C. KEPHART, University of Purdue. Ready in October

A treatment of the employment interview as a basic technique for evaluating, selecting and placing workers. The interview is considered not as an isolated process but as an integrated part of the entire selection procedure. The author takes the point of view that facts are relatively easy to get; the real problem is to know what facts are important, to know when they have been given, and to know what to do with them once they are gathered.

MEASUREMENT IN EDUCATION. An Introduction

By A. M. JORDAN, University of North Carolina. *McGraw-Hill Series in Education.* Ready in January

A well-written introductory text characterized by thorough scholarship, this book gives emphasis to the place of measuring instruments in the whole educative process. It considers the major areas of this process to be the definition of objectives, learning, and evaluation and appraisal. Many new types of measurement are included which point the way to proper emphasis in test construction.

THE DYNAMICS OF THE COUNSELING PROCESS

By EVERETT L. SHOSTROM, George Pepperdine College, Los Angeles; and LAWRENCE M. BRAMMER, Sacramento State College. 213 pages, $3.50

Based on the fundamental concept that the counselor's purpose should not be simply to solve the student's problems, but he should rather aim at facilitating the process of mature self-direction and providing a situation in which the student can help himself. The aim of this new text is to present a unified approach to educational-vocational-personal counseling.

Send for copies on approval

McGRAW-HILL BOOK COMPANY, Inc.

330 West 42nd Street New York 36, N.Y.